Models of the Unimodal and Multimodal Statistics of Adaptive Wavelet Packet Coefficients

Roberto Cossu, Ian Jermyn, Karen Brady, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Models of the Unimodal and Multimodal Statistics of Adaptive Wavelet Packet Coefficients

Roberto Cossu — Ian Jermyn — Karen Brady — Josiane Zerubia

N° 5122

Fevrier 2004

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--5122--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Models of the Unimodal and Multimodal Statistics of Adaptive Wavelet Packet Coefficients*

Roberto Cossu , Ian Jermyn , Karen Brady , Josiane Zerubia

Thème 3 —Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 5122 —Fevrier 2004 —43 pages

Abstract: In recent work, it was noted that although the subband histograms for standard wavelet coefficients take on a generalized Gaussian form, this is no longer true for wavelet packet bases adapted to a given texture. Instead, three types of subband statistics are observed: Gaussian, generalized Gaussian, and most interestingly, in some subbands, multimodal histograms with no mode at zero. As will be demonstrated in this report, these latter subbands are closely linked to the structure of the texture, and are thus likely to be important for many applications in which texture plays a role. Motivated by these observations, we extend the approach to texture modelling proposed by Brady et al. [4] to include these subbands. We relax the Gaussian assumption to include generalized Gaussians and constrained Gaussian mixtures. We use a Bayesian methodology, finding MAP estimates for the adaptive basis, for subband model selection, and for subband model parameters. Results confirm the effectiveness of the proposed approach, and highlight the importance of multimodal subbands for texture discrimination and modelling.

Key-words: wavelet packet, adaptive, multimodal, bimodal, statistics, texture, probabilistic, Bayesian

* This work was partially funded by EU project Moumir, HP-99-108, www.moumir.org. The authors would like to thank the IGN and the IFN for providing the remote sensing images.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Modèles des Statistiques Unimodales et Multimodales des Coefficients de Paquets d'Ondelettes Adaptatifs

Résumé : De récents travaux ont montré que bien que les histogrammes de sous-bandes pour les coefficients d'ondelettes standards ont une forme de gaussienne généralisée, ce n'est plus vrai pour les bases de paquets d'ondelettes adaptés à une certaine texture. Trois types de statistiques sont alors observés pour les sous-bandes: gaussienne, gaussienne generalisée et dans certaines sous-bandes des histogrammes multimodaux sans mode en zéro. Dans ce rapport, nous démontrons que ces sous-bandes sont étroitement liées à la structure de la texture et sont ainsi primordiales dans les applications dans lesquelles la texture joue un rôle important. Fort de ces observations, nous étendons l'approche de modélisation de textures proposée par Brady et al. [4] en incluant ces sous-bandes. Nous modifions l'hypothèse gaussienne pour inclure les gaussiennes généralisées et les mixtures de gaussiennes contraintes. Nous utilisons une méthodologie bayésienne, définissant des estimateurs MAP pour la base adaptative, pour la sélection du modèle de la sous-bande et pour les paramètres de ce modèle. Les résultats confirment l'efficacité de la méthode proposée et soulignent l'importance des sous-bandes multimodales pour la discrimination et la modélisation de textures.

Mots-clés : paquet d'ondelettes, adaptatif, multimodal, bimodal, statistiques, texture, probabiliste, bayesien

Contents

1 Introduction

The goal in image processing and computer vision is almost always to infer from the image information about the scene that generated that image. The nature of this information varies from application to application, and is captured by the semantics of the application: the set of statements one would like to be able to make about the images concerned. These inferences will be made using probability theory, and this leads to the emergence of two probability distributions of equal importance: a prior distribution on the space of statements one would like to make, which is frequently a model of the physical scene, though it ain't necessarily so; and a model function giving, for each possible statement, the probability of each image given that the statement is true. Together these define, via Bayes' theorem, the probability of each statement given the data, which can then be used to make inferences and estimates.

To give an example: in the case of remote sensing images, the elements of a set of statements of particular importance concern land cover types. One would like to be able to make statements of the form 'the land area corresponding to such-and-such region in the image domain is covered by such-and-such a land cover type'. The list of possible land cover types is itself application dependent, but typical examples are 'forest' and 'urban area'. Following the recipe of the first paragraph, two things are needed. First, the prior probability of any given set of statements must be defined. Since the Boolean algebra generated by the statements has natural anti-atoms, conjunctions of statements concerning regions that together partition the image domain, it suffices to define a probability distribution on these anti-atoms, the rest following by marginalization. Second, the probability of any image given a set of statements must be defined, although again it suffices to give the probability of an image given an anti-atom. Under two assumptions concerning the probability of the image, to wit, that image values in the region corresponding to a particular land cover type depend neither on image values in regions corresponding to different land cover types, nor on the identity of those land cover types, the probability distribution factorizes so that the quantity of primary concern is the probability of the image in a region corresponding to a single land cover type. It is on these probability distributions that this report focuses its attention.

1.1 Texture and image statistics

In remote sensing images at a range of resolutions, the images corresponding to particular land cover types may be characterized by the word 'texture'. Although ill-defined, this word captures the idea of 'regularity with variation' that applies to images of forest, ploughed field and so on. The same is true of the images that arise in many other domains. As a consequence, texture description and analysis have long played an important role in remote sensing applications [18], as well as in other image processing areas, for example, medicine [13], and industrial quality control [10, 7].

In one sense, texture is simple: it has a repetitive nature that means that probabilistic models describing it have simple properties compared to those that would be required to describe images of human beings, for example. On the other hand, it is in the nature of texture to possess spatial dependencies with a relatively long range, and this makes modelling it much harder than modelling white noise, for two reasons. First, the spatial dependencies must be incorporated into the model. Second, since in applications, as described above, we typically want the probability of the image in a finite region, we must in principle deal with the behaviour of these dependencies at the border of the region.

This need for spatial dependence coupled with spatial localization within regions means that wavelets are a natural tool to use in modelling texture. They are spatially extended, thus facilitating the description of spatial dependencies, but unlike the Fourier basis they are also spatially localized. They have been much used for texture description [15, 16, 19, 6].

Early models assumed that the wavelet coefficients were independent, and concentrated on characterizing the statistics of subbands accurately. It was rapidly discovered that the subband histograms of many classes of images (especially 'natural images') assumed leptokurtotic forms. These forms have been

modelled by Laplacian distributions, generalized Gaussians [16, 11], and by mixtures of two Gaussians, one of small and one of large variance, corresponding roughly to flat, slightly noisy areas, separated by a relatively small number of edges [8].

More sophisticated models introduced dependence between wavelet coefficients, the most well-known being Markov tree [1] and hidden Markov tree [6, 8] models, in which wavelet coefficients at a given scale depend on those at the next coarser scale, either directly or via hidden states. These tree models capture dependencies between scales well, and are thus good for describing edge structures, which persist across scale. Indeed, Bonet and Viola [1] deliberately set out to describe textures formed by edges, while Crouse et al. [8] set out to improve the description of natural images by encouraging edges to persist across scale. Choi and Baraniuk [6] then applied the models to texture description.

Although for certain classes of texture, the existence of edges is important, a more common feature is the existence of significant periodicities that run throughout the texture. Indeed even when edges are important, it is usually the fact that these edges appear in a regular structure that characterizes the texture itself. The dependencies responsible for such periodicities are difficult to capture within models based on standard wavelet bases. Standard bases couple interaction range and frequency, so that independent wavelets can describe long-range interactions only as taking place at low frequencies. What is needed are dependencies between the wavelet coefficients within a given scale. This rules out tree models, which, while good at capturing the interscale dependencies implied by edges, are not good at capturing the intrascale dependencies required here. Instead, what is needed are models that explicitly represent the dependencies within scales, and indeed such models were mentioned as an alternative to trees by Crouse et al. [8]. They were not much discussed, however, perhaps because the resulting models look like a collection of (possibly hidden) Markov random field models, one in each subband or at each scale, and are thus difficult to treat efficiently.

An alternative family of models, whose members are capable of capturing long-range dependencies in narrow frequency bands while maintaining tractability, was proposed by Brady et al. [2, 3, 4]. In these models, dependencies are captured not by explicit interactions between standard wavelet coefficients, but by the use of wavelet packet bases *adapted* to the dependencies present in a texture. The wavelet packet coefficients themselves are independent, thus ensuring the tractability of the models, while the relation between the wavelet packet basis and the standard wavelet basis means that the latter coefficients are dependent. The models in the family are Gaussian, with inverse covariances lying in the large set of operators diagonal in *some* wavelet packet basis. A Bayesian methodology is used to compute an exact MAP estimate of the covariance from samples of the texture class concerned, which has the side effect of specifying a preferred wavelet packet basis for each class, in which the covariance is diagonal.

The Gaussian assumption appears to go against the above-mentioned observation that in standard wavelet bases, the subband histograms of many classes of images are leptokurtotic. There are two problems with this argument, First, most of the observations reporting this phenomenon were performed on 'natural' images, rather than on individual texture classes. The natural image histograms are thus mixtures of histograms derived from more specific distributions, and one need not expect the components of a mixture to share the same form as the mixture itself. Second, even if the marginal distributions of a coherent texture were leptokurtotic, the marginal distributions derived from products of generalized Gaussians (or Gaussian mixtures) modelling independent subbands do not preserve their shape under changes of basis, and there is thus no reason to expect that the subband histograms of wavelet packet bases will necessarily be leptokurtotic. This seems especially true when the wavelet packet basis is adapted to the texture using a Gaussian model. Thus in the absence of further knowledge, the Gaussian distribution is a good 'minimal' distribution to use, for many reasons [14], even though it cannot produce leptokurtotic marginal distributions.

The failure of leptokurtosis for adaptive wavelet packet bases was confirmed by Brady et al. [2]. Despite the use of a Gaussian model, many of the adaptive subband histograms remain generalized Gaussians; many of the remainder, however, appear Gaussian. Most interestingly of all, in some subbands

a completely different behaviour is observed: the subband histograms are bi- or more generally multimodal. These subbands are usually those that are narrow in frequency content; the presence in the subband histograms of maxima at non-zero coefficient values thus indicates the likely presence of approximate periodicities running throughout images belonging to that texture class, with frequencies in the support of these subbands. These multimodal subbands thus represent the characteristic structure of a texture class.

It is worthwhile emphasizing both the intuitive nature of this result, and its importance for texture modelling. Existing models of texture classes are composed of mixtures of products over a basis of unimodal distributions whose mode is at zero (we include the case of mixtures with one component). The only exception is in the modelling of the scaling coefficients, or other additive constants, which may have non-zero mean and mode. Note that this applies not only to independent subband models, but to hidden Markov tree models as well. The result is that the most probable image in these texture models is in fact flat, that is, without texture. This clearly contradicts our prior knowledge of the type of images to be expected of 'forest', for example; in general it seems that the one thing one can say for sure about images of textures is that they are not flat. The distributions describing them thus possess modes away from zero.

Another way of seeing the same point is to consider the argument normally given for the leptokurtotic distributions found for standard wavelet coefficients. This is that the leptokurtosis is created by the combination of a large number of small wavelet coefficients, essentially corresponding to noise, and a small number of large coefficients corresponding to edges. The fact that edges are sparse in 'natural' images, leads to sparse wavelet subbands, and leptokurtosis. However, in a texture that possesses a periodicity that runs throughout its spatial extension, and that therefore is not sparse, there is no reason to suppose that one will find a large number of small coefficients if the wavelets themselves are sufficiently focused on the frequency band of the periodicity. If the wavelets are not so focused, however, the sparseness of frequencies other than that of the periodicity will swamp its effect.

Clearly, this empirical result is important for accurate texture modelling, and hence for applications in which such modelling plays a role. The presence of multimodal subbands should provide a powerful distinguishing feature for texture classes. Motivated then by all the above, in this report we do the following:

1. In section 2, we review the approach of Brady et al. [4] and **demonstrate the existence of multimodal subbands** by applying this approach to a texture from the Brodatz album.

2. In section 3, we make precise the above arguments concerning the **inevitability of multimodal behaviour**, using a variety of simple texture models constructed in the Fourier domain. These confirm intuition, and also show that **multimodality is inextricably linked to the characteristic structures in a texture**. We also demonstrate empirically the effect on the wavelet packet coefficient histograms of adding a periodicity to an image. Along with the evidence from the previous section, and that to be shown in section 5, this motivates the modelling work in the subsequent section.

3. In section 4, we describe **new image models that capture the newly observed behaviour**. We relax the Gaussian assumption to consider three possible models for each adaptive subband: Gaussian, generalized Gaussian, and a constrained mixture of three Gaussians. We extend the Bayesian methodology used by Brady et al. [4] to compute MAP estimates for the adaptive basis, for the choice of subband models, and for the subband model parameters.

4. In section 5, we first discuss the sensitivity of the models to the choice of two parameters that for the moment are fixed by hand. We then present, on textures from the Brodatz album and from remote sensing images, **results demonstrating the accuracy of the modelling**. Finally, we demonstrate the **usefulness of multimodal subbands for texture classification** even when used on their own, via a closer examination of the statistics of such subbands.

5. In section 6, we conclude briefly.

2 Texture: an Adaptive Probabilistic Approach

In this section, we briefly review the approach of Brady et al. [4].

A texture is modelled by a probability distribution that expresses the probability of an image given that it is an image of that texture. Since planar texture is, by its nature, infinitely extendable, these probability distributions are defined on the space of infinite images. However, applications necessitate the probability of an arbitrarily-shaped region of the texture. If the distribution on infinite images is Gaussian, the probability distribution of a region is also Gaussian, but with a covariance that is a complicated region-dependent function of the original, infinite-image covariance. However, a restriction on the infinite-image covariances allows the resulting region covariances to be diagonalized using a wavelet packet basis that depends on the form of the covariance. Effectively, the wavelet packet basis is chosen so as best to capture the correlations that exist in the texture, while remaining independent.

In more detail, textures are modelled by a distribution $\Pr(\phi|K)$, where ϕ is an image in the space of infinite images Φ_∞, which we will take to be a d-dimensional torus T^d of size $L = \infty$. K represents all the prior knowledge necessary to determine the distribution, for example its parameters. A Gaussian model is then

$$\Pr(\phi|K) = |F/\pi|^{1/2} e^{-\langle \phi-\nu|F|\phi-\nu\rangle} , \tag{1}$$

where: $\langle\phi|\psi\rangle$ denotes the inner product of two images; F is the inverse covariance operator; $F|\phi\rangle$ means F acting on $|\phi\rangle$; $\nu \in \Phi_\infty$ is the mean; and $|F|$ denotes the determinant of F.

Let us consider a region R of the image domain, and its complement $\bar{R}$. Marginalizing over the values of pixels in $\bar{R}$ gives the following probability measure for the image ϕ_R on the region R:

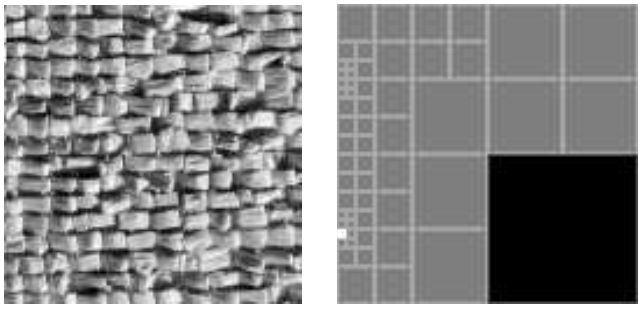
$$\Pr(\phi_R|K) = |G_R/\pi|^{1/2} e^{-\langle \phi_R-\nu_R|G_R|\phi_R-\nu_R\rangle} , \tag{2}$$

where $G_R = F_{RR} - F_{R\bar{R}}F_{\bar{R}\bar{R}}^{-1}F_{\bar{R}R}$ and F_{RR}, $F_{R\bar{R}}$, $F_{\bar{R}\bar{R}}$, and $F_{\bar{R}R}$ correlate R with itself, R with $\bar{R}$, $\bar{R}$ with itself, and $\bar{R}$ with R, respectively. The computational complexity of evaluating this operator requires G_R to be diagonalized. For general F, this is a difficult task. However, the requirement that the distribution be translation invariant helps because it means that F is diagonal in the Fourier basis, and thus is characterized by a function f on the Fourier domain. By restricting attention to the set of functions that are piecewise constant on at least one dyadic partition of the Fourier domain, we ensure that the covariance F is diagonalized not only by the Fourier basis, but also, approximately, by a wavelet packet basis defined by the dyadic partition on which f is piecewise constant. The distribution for the image is then given by

$$\Pr(\phi|(T,f),K) = \prod_{t\in T} |f_t/\pi|^{N_t/2} e^{-f_t \sum_{i\in t}(w_{t,i}-\nu_t)^2} , \tag{3}$$

where: t is an element of the dyadic partition T (*i.e.* a subband); f_t is the (constant) value of the function f on t; ν_t is the (constant) value of the mean on t (in fact, the mean will be zero except for the scaling coefficient subband); N_t is the number of wavelet packet coefficients in t; and $w_{t,i}$ are the wavelet packet coefficients in t, indexed by i. (The range of i varies from subband to subband, but we do not notate this for reasons of clarity.)

The use of this same probability distribution to describe the image ϕ_R on a periodic region R of dyadic size (*i.e.* a dyadic torus) is unproblematic, since the wavelet packets form an exact basis for such a space. (Indeed, this is one of the points of using wavelets.) For a non-periodic dyadic region, the errors are small. For an arbitrarily shaped region, however, the wavelet packets contained within the region provide



a b

Figure 1: Raffia texture (a) and its adaptive decomposition (b). Top-left: low-low frequencies; bottom-right: high-high frequencies. Subband t_1 and t_2 are shown in black and white, respectively.

only an approximate basis. While this case is important for image segmentation and classification, it does not greatly concern us in the present work. Various ways of dealing with this problem are discussed by Brady et al. [2, 3, 4].

The parameters of the model (T, f and ν) are taken to be equal to their MAP estimates based on a set of training data from the texture class being modelled. Thus both T and f depend on the texture class. These MAP estimates can be calculated using an efficient depth-first search algorithm. This algorithm, and the prior probabilities used by Brady et al. [4], are special cases of the algorithm and prior probabilities described in section 4, where several models for each subband are considered.

Once the texture parameters are estimated, the model can be used in a number of different applications in which the description of textured images is required. For instance, Brady et al. [4] address image classification. In this report, we focus our attention on the statistical properties of the wavelet packet coefficients in different subbands.

2.1 The multimodal statistics of adaptive wavelet packet coefficients

Having generated a wavelet packet basis, T, for each texture class considered, it is natural to examine the empirical statistics of the resulting wavelet packet coefficients. Non-Gaussian behaviour is observed in two classes of subband. In particularly sparse subbands, the histograms tend to be leptokurtotic, as would be expected. In figure 1, the Raffia texture from the Brodatz album is shown alongside its adaptive wavelet packet decomposition. The histogram of the wavelet packet coefficients in the subband t_1, shown in black in figure 1, is shown in figure 2. Clearly this is modelled better by a generalized Gaussian with a shape factor smaller than 2 than by a Gaussian.

The second and more important case of non-Gaussianity occurs in subbands with narrow frequency content. The histogram of the wavelet packet coefficients in the subband t_2, shown in white in figure 1, is shown in figure 3. Note that the histogram has a clear bimodal form. It was observations such as this, more examples of which will be given in section 5, that prompted the empirical and theoretical investigation and modelling described in this report.

3 Multimodality: Theory

Having noted the existence of subbands with multimodal histograms, we now attempt to give an explanation for this behaviour. In order to predict histograms of wavelet coefficients, we need an image model,

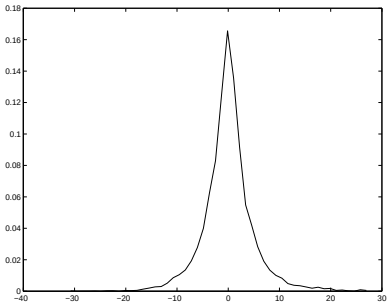


Figure 2: Wavelet packet coefficient histogram for the subband t_1 indicated in figure 1.

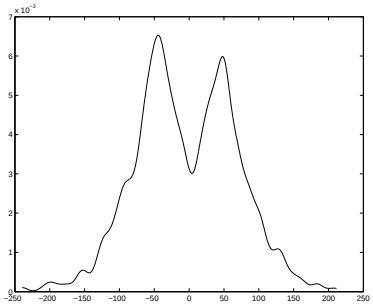


Figure 3: Wavelet packet coefficient histogram in the subband t_2 indicated in figure 1

that is, a probability distribution $\Pr(\phi)$ on the space of images Φ. (In this section, we again assume that images are functions on a d-dimensional, discrete or continuous torus T^d, but now of finite dyadic size L.) Given such a model, we can predict the histogram of wavelet coefficients we will observe by computing, for example, the mean histogram under this distribution. In the following, we take various image models and predict the mean histograms of the wavelet coefficients, showing how multimodal histograms can arise, and linking them with the characteristic structure in a texture, *i.e.* its principal periodicities. We do not use wavelet based models directly, since this would seem to be begging the question. Rather we use various reasonable models of textures constructed in the Fourier domain.

3.1 Mean histograms

Given a set of images $\{\phi_n : n \in N\}$, the histogram of the wavelet coefficients in a given subband is

$$H[\{\phi_n\}](\omega) = \sum_n \sum_i \delta(\omega, W_i(\phi_n)) \ ,$$

where i indexes the individual wavelet coefficients $W_i(\phi)$ in the subband. The mean histogram for that subband is then given by

$$\langle H(\omega) \rangle = \int_{\Phi^n} \Pr(\{\phi_n\}) \, H[\{\phi_n\}](\omega) \ .$$

If we assume that the images in the set are independent, then the effect of multiple images is simply to multiply the mean histogram by their number (although the variance of the distribution for H will go down with increasing numbers of images). We therefore consider only one image, ϕ, so that

$$\langle H(\omega) \rangle = \int_\Phi \Pr(\phi) \sum_i \delta(\omega, W_i(\phi)) = \sum_i \int_\Phi \Pr(\phi) \, \delta(\omega, W_i(\phi)) \ .$$

Now note that the distribution $\Pr(\phi)$ must be translation invariant, since we do not know how the image will be translated when we see it, and that the W_i are related by translation. Thus

$$\int_\Phi \Pr(\phi) \, \delta(\omega, W_i(\phi)) = \int_\Phi \Pr(\phi) \, \delta(\omega, W_0(T_i \phi)) = \int_\Phi \Pr(T_i^{-1}\phi) \, \delta(\omega, W_0(\phi)) = \int_\Phi \Pr(\phi) \, \delta(\omega, W_0(\phi)) \ .$$

Here T_i is the translation relating the zeroth and the i^{th} wavelet coefficient. Writing W_0 as W, we then have that

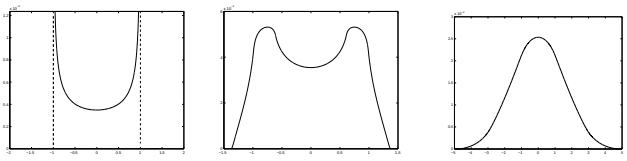
$$\langle H(\omega) \rangle = Z^{-1} \int_\Phi \Pr(\phi) \, \delta(\omega, W(\phi)) \ ,$$

where Z is a normalization constant, to be dropped hereafter.

Clearly the mean histogram only depends on the marginal probability of an image in the spatial and frequency support of the wavelet coefficient W. We will work primarily in the Fourier domain, in which case

$$W(\phi) = \sum_{k \in K} w_k^\dagger \phi_k = \sum_{k \in K^+} \tilde{\phi}_k = \sum_{k \in K^+} A_k \cos(\theta_k) \ .$$

Here w_k (resp. ϕ_k) is the Fourier coefficient of W (resp. ϕ) at frequency k, while K (resp. K^+) is the frequency support of the wavelet coefficient (resp. contained in an arbitrarily defined 'positive' half of



a b c

Figure 4: Histograms corresponding to different image models: a) one non-zero Fourier coefficient with fixed amplitude and uniformly distributed phase; b) three non-zero coefficients with fixed amplitudes of 1, 0.01, and 0.01, and independently and uniformly distributed phases; c) five non-zero coefficients of equal amplitude and independently and uniformly distributed phases.

the Fourier domain). In the second and third equations, we have set all the $w_k = 1/2$, giving the ideal bandpass filter. A_k and θ_k are the amplitude and phase of ϕ_k respectively, while $\tilde{\phi}_k$ is its real part. Thus

$$\langle H(\omega)\rangle = \int \Pr(\tilde{\phi})\, \delta(\omega, \sum_k \tilde{\phi}_k) = \int \Pr(A,\theta)\, \delta(\omega, \sum_k A_k \cos(\theta_k)) \;, \tag{4}$$

We note that if the $\tilde{\phi}_k$ are all independent, then the full mean histogram is given by the convolution of their individual mean histograms.

3.2 Single sinusoid

Let us suppose that all the A_k are zero, except for $A_0 = A_{k_0}$, which is equal to a fixed value A, while the θ_k are all uniformly distributed on S^1:

$$\Pr(A,\theta) = DA D\theta\, \delta(A_0, A) \left[\prod_{k \in K^+ \setminus k_0} \delta(A_k) \right] \;,$$

where $DA = \prod_k dA_k$ and $D\theta = \prod_k (d\theta_k/2\pi)$. The set of images thus consists of a sinusoid of fixed frequency and amplitude, but with unknown translation. In this case, equation (4) is easy to calculate. The result is

$$H_A(\omega) \triangleq \langle H(\omega)\rangle = \int_{-\pi}^{\pi} \frac{d\theta_0}{2\pi}\, \delta(\omega, A\cos(\theta_0)) = \begin{cases} \frac{1}{\pi}(A^2 - \omega^2)^{-1/2} & \text{if } -A \le \omega \le A \;, \\ 0 & \text{otherwise .} \end{cases} \tag{5}$$

Figure 4(a) shows a plot of this function. The intuitive reason for the result is clear: a sinusoid 'spends more of its time' near its maximum value than it does near zero.

If we relax the constraint on A_0, for example by replacing the delta function in equation (5) by a narrow Gaussian (to avoid problems with the positivity of the A_0), we still find that the mean histogram is bimodal: it is a blurred version of figure 4(a). We get a similar result if we use a Gamma distribution for A_0^2, which produces a bimodal mean histogram for all values of the index greater than $3/2$, although the bimodality occurs on an increasingly small scale as the index approaches $3/2$. When the index is exactly $3/2$, the distribution for A_0 becomes Rayleigh, at which point, as is well known, the distributions for ϕ_k and $\tilde{\phi}_k$, and hence the mean histogram of ω, become zero-mean Gaussians, and bimodality is lost. Similar gradual loss of bimodality is observed by adding complex-valued noise of increasing variance to ϕ_k. Bimodality is thus intrinsically linked to the degree of certainty in the Fourier amplitude.

3.3 Multiple independent sinusoids

We consider now the case in which more than one of the A_k is non-zero. First, let them all have fixed values B_k, while the phases θ_k are all independently and uniformly distributed, as before:

$$\Pr(A, \theta) = DAD\theta\, \delta(A, B) \ . \tag{6}$$

The A_k integrals in equation (4) are again trivially performed, giving

$$H_{\{B_k\}}(\omega) = \int \left[\prod_k \frac{d\theta_k}{2\pi} \right] \delta(\omega, \sum_k B_k \cos(\theta_k)) \ .$$

The mean histogram is therefore the convolution of a number of mean histograms $H_{B_k}(\omega)$ of the form in equation (5). If among the non-zero B_k values, one of them has an energy significantly higher than the others, then the effect of this convolution is to round the bimodal function in equation (5). The small B_k values have an effect somewhat like the noise discussed at the end of the previous section. The result is shown in figure 4(b).

On the other hand, if the B_k all have approximately the same value, then the individual mean histograms are all the same, and thus, as the number of non-zero B_k values increases, their convolution tends towards a Gaussian, as in the central limit theorem. The double peak disappears using only a few values. The resulting mean histogram is shown in figure4(c). The result clearly holds whenever the $\tilde{\phi}_k$ are iid.

To validate the above, we carried out the following numerical experiment. The values B_k were computed from

$$B_{k_n} = \exp\{-(k_n - k_0)^2 / 2\sigma^2\} \ ,$$

where $\{k_n = k_0 + n\delta k : \ n \in [-N, \ldots, N]\}$ were a number of equally spaced frequencies centred on k_0, and σ, when it is small enough, effectively controls the number of non-zero B_k. We then sampled, for each k_n, many random phases $\theta_{n,i}$, and computed the histograms of the resulting values of $\sum_n B_{k_n} \cos(\theta_{n,i})$. In figure 5 the resulting histograms are plotted for varying values of σ. For small σ, that is with only a few non-zero B_{k_n}, and with B_{k_0} much larger than $B_{k_{\pm 1}}$, the histograms are strongly bimodal. As σ is increased, the histogram is smoothed, until for large enough sigma, the bimodality is destroyed and the histogram becomes unimodal. This is visualized in 3D in figure 6.

3.4 Single translated signal

In the image models considered so far, the $\tilde{\phi}_k$ were independent. We now consider a case in which the phases are dependent, while the amplitudes remain fixed. In particular, we consider a distribution whose support is all translations of a given signal:

$$\Pr(A, \theta) = DAD\theta\, \delta(A, B) \int_{T^d} \frac{dl}{L} \prod_{k \in K^+} \delta(\theta_k, \theta_{k,0} + kl) \ .$$

As a consequence of the dependence between the phases, the histogram can no longer be computed as a convolution. It is not easy to obtain general analytical results, but numerically it is simple. We proceeded as follows. In order to construct the A_k and $\theta_{k,0}$, we chose a finite-length discrete sinusoid for which the signal length was not a multiple of the period. The discrete Fourier transform of this signal presents a number of non-zero coefficients with fixed phase relationships. If these phase relationships are not preserved, then the structure of the 'texture' is lost. We then created a large number of shifted versions of

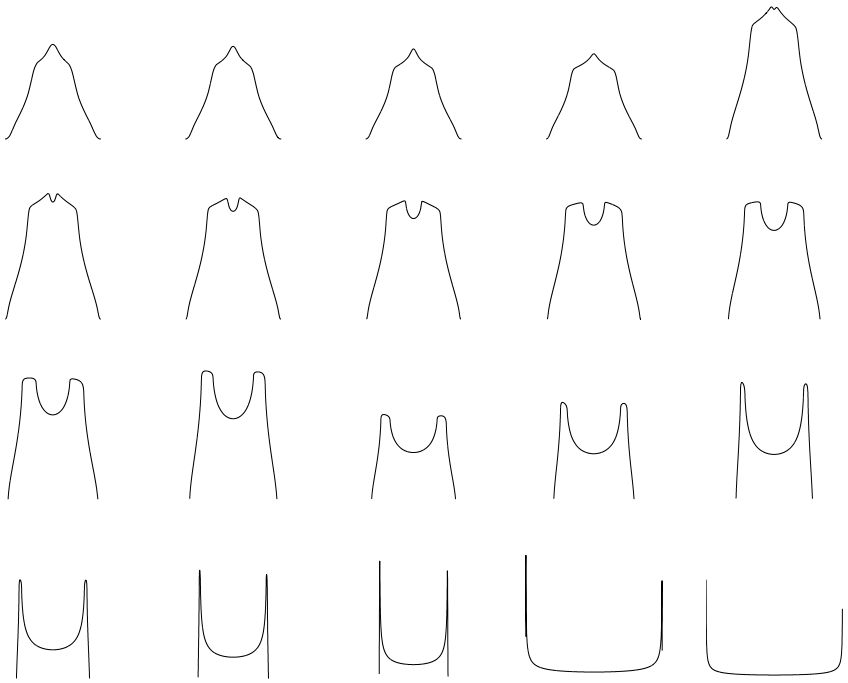


Figure 5: Histograms of wavelet coefficients for a signal whose Fourier coefficient amplitudes at neighbouring frequencies take on a Gaussian shape plotted against σ.

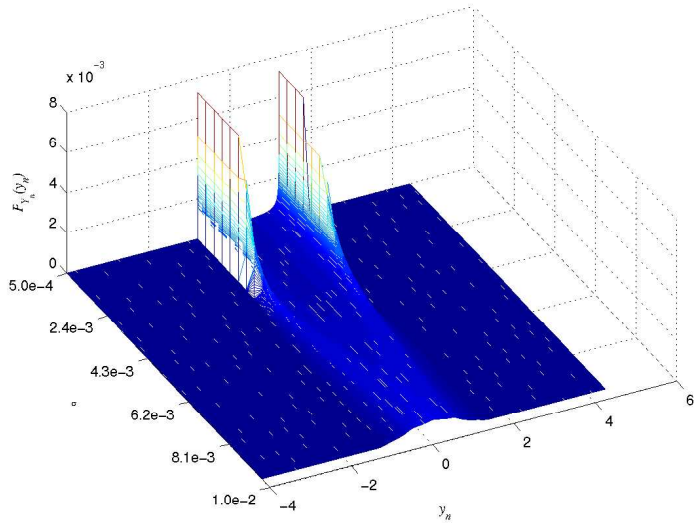
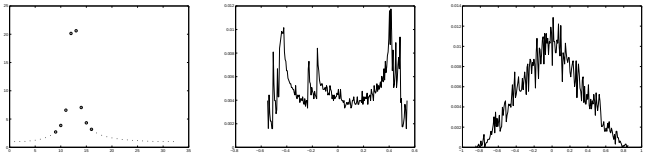


Figure 6: 3D representation of the plots in figure 5.

the signal by uniformly sampling values of t, each shifted version then being filtered and downsampled (using an ideal bandpass filter). We then computed the histogram of the resulting signals.

Figure 7(a) shows the DFT of a discrete sinusoid of length 64 and frequency 0.18. We considered 6400 equally spaced translations of the signal, obtaining 6400 different values of the phase for each DFT coefficient. At this point we filtered the signal, extracting the coefficients with wavenumbers from 9 to 16. The resulting histogram is plotted in figure 7(b). As one can see, despite considering several non-zero coefficients, the histogram takes on a bimodal form. The preservation of the bimodality is due to the relations that exist between the phases, as can be seen by looking at figure 7(c), which shows the histogram of a signal whose DFT amplitudes are the same as before but whose phases were sampled independently. The histogram is unimodal, in agreement with the results shown in figures 5 and 6.



a b c

Figure 7: Absolute values of the DFT of a finite-length discrete sinusoidal signal whose length is not a multiple of its frequency (a). Histogram obtained when: (b) retaining phase information; (c) discarding phase information

3.5 Simulated bimodal subbands

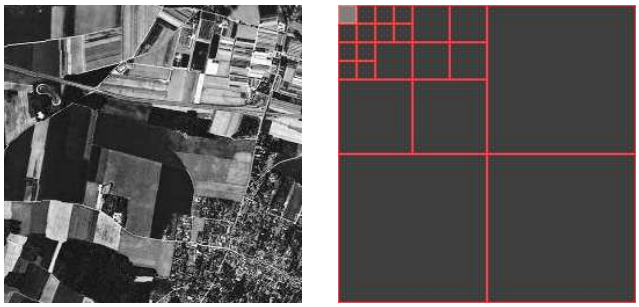
The theory developed in the previous subsections indicates that bimodal histograms appear when

- The wavelet filter is tuned to a range of frequencies that corresponds to a peak in the Fourier transform, *i.e.* a significant periodicity;
- The filter bandwidth is narrow enough to isolate such a feature from the background.

In this subsection, we illustrate these points experimentally in the following way. Consider an image ϕ (shown on the left of figure 8) in which no particular periodicities are discernible. The adaptive wavelet decomposition of section 4 applied to this image results in the decomposition shown on the right of figure 8. The histograms of the wavelet coefficients in the different subbands of the decomposition are plotted in figure 9. As can be seen, the decomposition found is similar to that of a standard wavelet transform, and none of the histograms presents significant bimodal behaviour. In fact, all the subbands are accurately modelled by GG curves. Now let us considered a new image ψ, obtained as follows. Let $\hat{\phi}$ be the Fourier transform of ϕ, and let ψ be the image whose DFT $\hat{\psi}$ is given by

$$\hat{\psi}(k) = \begin{cases} A & \text{for } k = \pm K \ , \\ 0 & \text{for } \|k \pm K\|_{l_\infty} < \epsilon \text{ and } k \neq \pm K \ , \\ \hat{\phi} & \text{elsewhere .} \end{cases} \tag{7}$$

The new image ψ is similar to the original image ϕ, except that a periodicity is added (a spike of height A of frequency K). The zeros are inserted in order to make it easy to isolate this periodicity from the background. Figures 10 and 11 show the image ψ and its DFT respectively. Figure 12 shows the result of applying the adaptive wavelet decomposition of section 4 to this new image. In contrast to the previous decomposition, the decomposition is deeper around the artificially introduced periodicity. The



a b

Figure 8: Image used for simulating bimodal distributions (a) and its adaptive decomposition (b).

periodicity is isolated and four bimodal subbands result. The histograms of the wavelet coefficients in the different subbands are plotted in figure 13. For further illustration, figure 14 shows the detail images, the frequency support, and the wavelet coefficient histograms for the four bimodal subbands. It is worth noting that

- All the detail images present sinusoidal behaviour;
- Their frequency supports surround the frequency K.

This experiment shows two things. First, it confirms that bimodality appears when the subband isolates a strong periodicity in the image under analysis. Second, it confirms that the proposed approach is capable of identifying, isolating, and modelling the main periodicity in a given texture.

4 Multimodality: Modelling

In section 3, we produced theoretical explanations for the presence of bimodal histograms in wavelet coefficient subbands using image models based on the Fourier domain. The purpose was to explain how these histograms might arise without 'rigging' the models, as would have been inevitable had we constructed them in the wavelet domain directly. However, while the Fourier basis is convenient for intuition, it is inconvenient in many applications. We are frequently in need of the probability of the data in an image region of an arbitrary shape, which is therefore not analysable using the Fourier basis. One of the reasons that wavelet bases are so popular is precisely that they provide good approximations to bases for regions of arbitrary shape, while remaining localized in the frequency domain. Indeed, it was the desire for a basis that was at the same time localized but still capable of describing the frequency content of texture accurately, that led to the use of adaptive wavelet packet bases by Brady et al. [4]. In this section, we will therefore take the bimodal histograms as read, and construct image models to describe this behaviour directly in the wavelet domain.

It is obvious that the Gaussian assumption needs to be relaxed. It is not that the Gaussian distribution is positively harmful; it is after all the maximum entropy distribution with given mean and variance, and as such assumes as little as possible about the wavelets in a subband given these values. It is clear however that it does not take advantage of the structure that is present in the bimodal subbands. In particular, it discards the information concerning the non-zero peak positions, which would appear to provide a valuable means of discriminating between textures.

Although the adaptive wavelet packet bases used by Brady et al. [2] were derived under a Gaussian assumption, the usefulness of the adaptive wavelet packet decomposition extends beyond this assumption.

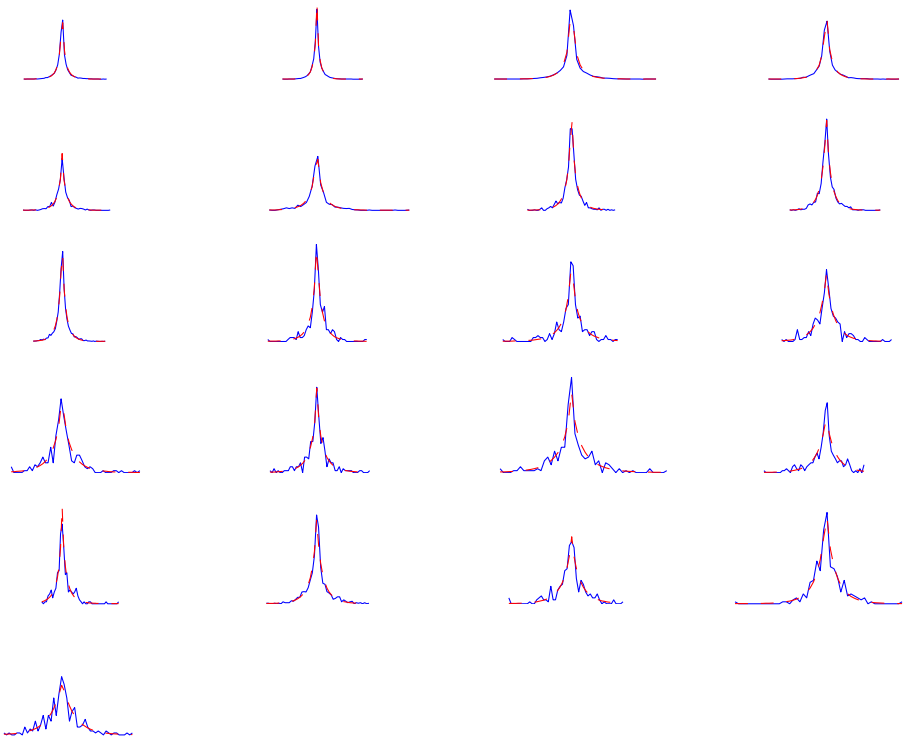
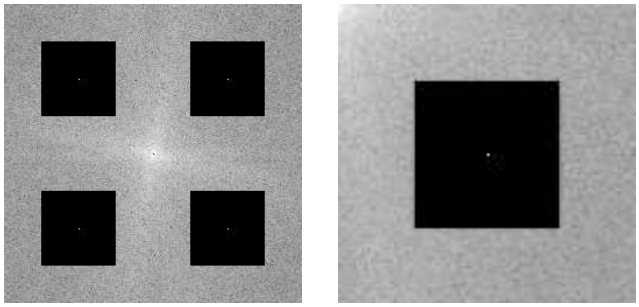


Figure 9: Wavelet coefficient histograms and fitted models for the subbands into which the image shown in figure 8 is decomposed. Observed histograms in blue (solid line); fitted models in red (dashed line).

Figure 10: Synthetic image created for simulating bimodal distributions. The periodic pattern corresponds to the artificially introduced sinusoidal component.



a b

Figure 11: Log display of the Fourier magnitude (a) and of the cropped Fourier magnitude (b) of the image in figure 10.

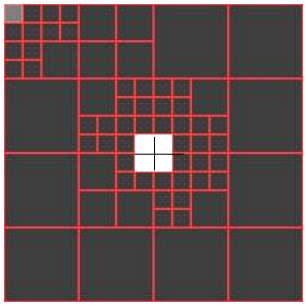


Figure 12: Adaptive decomposition obtained for the image shown in figure 10. Different colours correspond to the different models automatically selected within each subband: black, gray, and white represent the GG model, the G model, and the MoG model respectively.

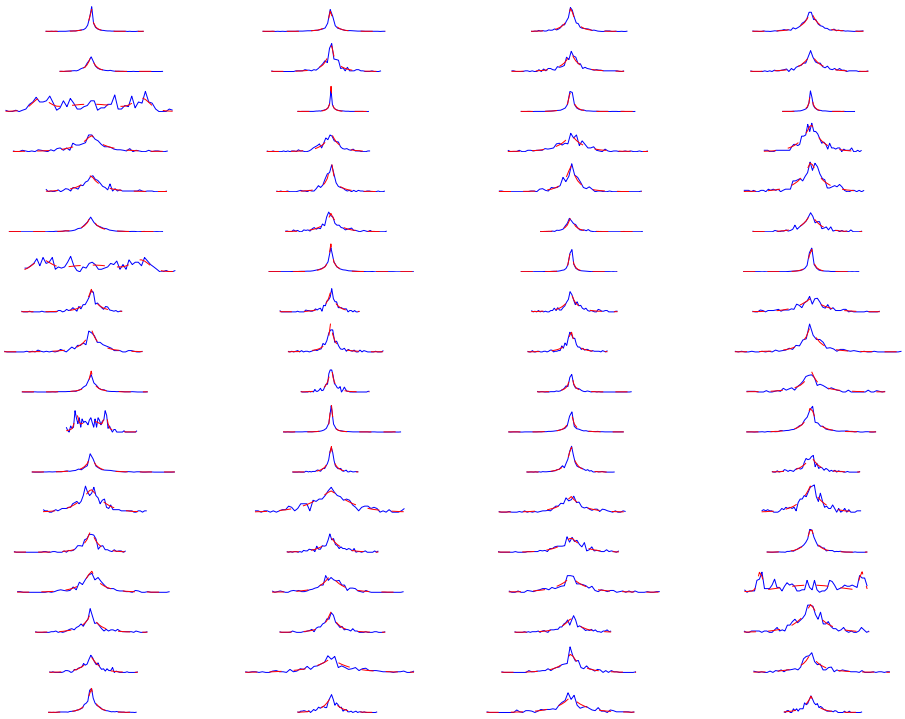


Figure 13: Wavelet coefficient histograms and fitted models for the subbands into which the image shown in figure 10 is decomposed. Observed histograms in blue (solid line); fitted models in red (dashed line).

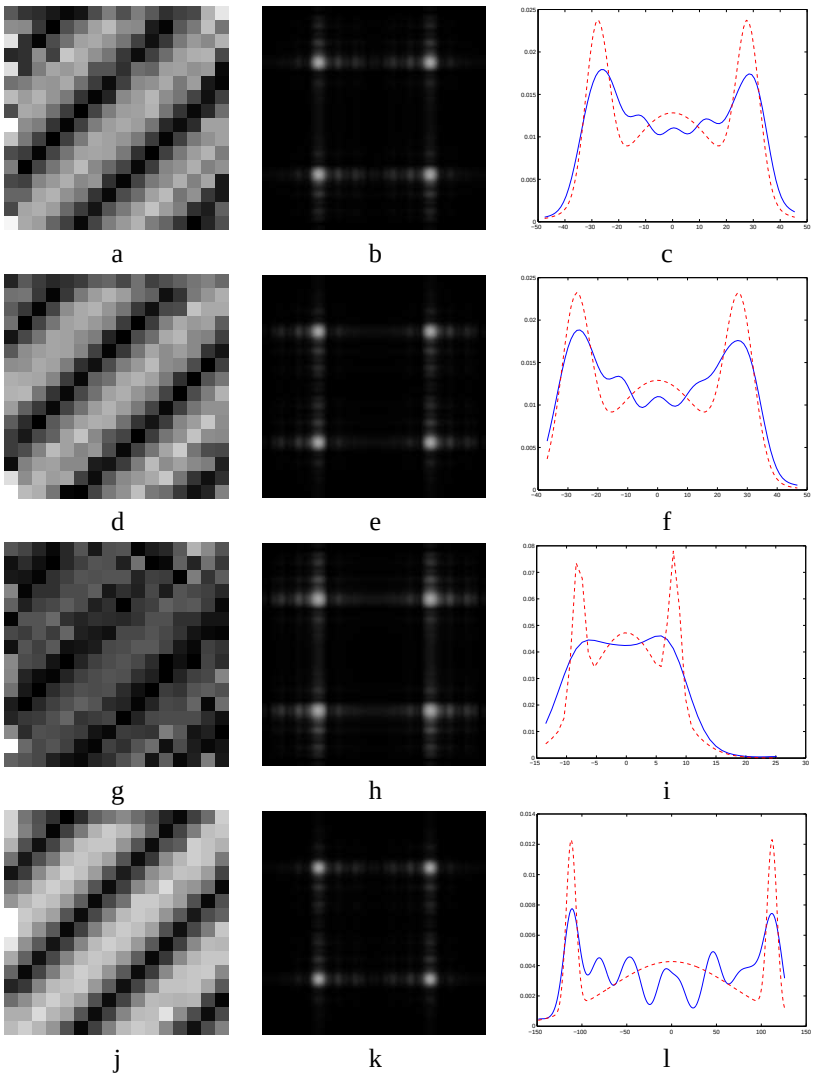


Figure 14: Analysis of the different bimodal subbands found in the decomposition of the image in figure 10. For each subband, the detail image, the frequency support, and the histogram are shown in the first, second, and third columns respectively. Observed histograms in blue (solid line); fitted models in red (dashed line).

It allows the analysis to focus on important regions of the Fourier domain, while maintaining spatial localization; and adaptivity means that it is reasonable to assume that the different subbands so identified are approximately independent of one another.[1]

We thus propose to create new image models by extending the adaptive probabilistic approach to include several models for the distributions of the wavelet coefficients in each subband, coupled with the assumption of subband independence. In each subband, we will suppose that the distribution is either Gaussian (G), generalized Gaussian (GG) or a constrained mixture of Gaussians (MoG). The image models that we will consider are thus parameterized by the following data:

1. a dyadic partition of one quadrant of the Fourier domain, T, which, given a mother wavelet, defines a wavelet packet basis;
2. a map μ from T to a set M of three models, $\{\mathrm{G}, \mathrm{GG}, \mathrm{MoG}\}$, giving the model used in each subband;
3. a map θ from T to the space of model parameters for each subband.

Note that dyadic partitions are in bijective correspondence with quadtrees with ordered child nodes (hereafter 'quadtrees'). We will use the language of partitions and trees interchangeably. Thus the terms 'partition' and 'tree', and 'element of a partition' and 'leaf node' are synonymous.

Since our aim is model building, the task facing us is to infer the values of the above parameters, given a number of examples of a particular class of textures. In the rest of this section, we explain how this is done.

4.1 MAP estimate: probabilities

We will estimate values of the above quantities using MAP estimation. That is, given a number of training images $\{\phi_n : n \in N\}$, we will calculate the probability

$$\Pr((\theta, \mu, T) | \{\phi_n\}, A, \beta) \propto \Pr(\{\phi_n\} | (\theta, \mu, T)) \Pr(\theta | \mu, T, A) \Pr(\mu | T) \Pr(T | \beta) \ . \tag{8}$$

where A and β are parameters to be described, and for the moment, assumed known. We will then choose those values of T, μ, and θ that maximize this probability.

The first factor on the right-hand side, the likelihood of the training data, is given by the product over N of the probabilities of the images individually, which in turn are given by (we suppress the index n for readability)

$$\Pr(\phi | (\theta, \mu, T)) = \prod_{t \in T} \Pr(w_t | \theta(t), \mu(t)) \ , \tag{9}$$

where w_t is the set of wavelet packet coefficients in the subband t. The probability in each subband is in turn given by the distribution corresponding to $\theta(t)$ and $\mu(t)$, which will be specified below.

The second factor, the probability of the model parameters given the model assignments to the tree, $\Pr(\theta | \mu, T)$, is also a product over the subbands. For each subband, all the parameters are assumed to be uniformly distributed over a very large range, A. Thus we have

$$\Pr(\theta | \mu, T, A) = \prod_{t \in T} A^{-\dim(\mu(t))} = \prod_{t \in T} e^{-\dim(\mu(t)) \ln(A)} \ , \tag{10}$$

where $\dim(m)$ is the dimension of the parameter space associated to the model $m \in M$. Note that this term penalizes model complexity. A is not determined automatically, but the results of model selection are not very sensitive to its value once an order of magnitude has been set.

[1] Note however that this does not mean that the standard wavelet coefficients are independent. Indeed one interpretation of the adaptive models, even in the Gaussian case, is as a way of encoding complex correlations between standard wavelet coefficients.

The third factor, $\Pr(\mu|T)$, is also simple. We assume that all models are equally likely *a priori*, with probability $1/3$, so that

$$\Pr(\mu|T) = \prod_{t\in T}(1/3) = e^{-|T|\ln(3)} ,$$

where $|T|$ is the cardinality of the partition T.

The final factor, $\Pr(T|\beta)$ is given by

$$\Pr(T|\beta) = Z^{-1}(\beta)e^{-\beta|T|} , \tag{11}$$

where β is a complexity penalty, and $Z(\beta)$ is a normalization factor. This prior probability penalizes bases with many subbands, and effectively regularizes the functions μ and θ viewed as functions of frequency. Note that both $\Pr(\theta|\mu, T, A)$ and $\Pr(\mu|T)$ also contribute to this penalty.

4.2 MAP estimate: algorithm

Note that equation (8) has the form

$$\begin{aligned}\Pr((\theta,\mu,T)|\{\phi_n\},A,\beta) &= \prod_{t\in T} e^{-E(w_t,\theta(t),\mu(t),t,A,\beta)} \\ &= e^{-L((\theta,\mu,T)|\{\phi_n\},A,\beta)} = e^{-\sum_{t\in T} E(w_t,\theta(t),\mu(t),t,A,\beta)} ,\end{aligned} \tag{12}$$

where E is defined by combining all the above equations. This form of the probability dramatically simplifies the computation of the MAP estimates, in the following way. For a given partition T, only the first two factors in equation (8) enter the estimates for μ and θ. Furthermore, the estimates of $\mu(t)$ and its parameters $\theta(t)$ for each element of T are functions only of the data w_t at that element, to be denoted $\mu^*(t, w_t)$ and $\theta^*(t, w_t)$ respectively, and are thus constant on the equivalence class of partitions that contain that element. The result is that the function E above becomes a function of t and w_t only, to be denoted $E^*(t)$, leading to a function $L^*(T) = \sum_{t\in T} E^*(t)$ (we suppress mention of w_t, A, and β).

To see the consequences of this, consider that the MAP estimate of T involves a search over the set $\mathcal{T}$ of all dyadic partitions (or equivalently quadtrees) down to some maximum depth. This set is enormous, and cannot be searched exhaustively. Now let $\oplus$ be the quaternary operator that maps four quadtrees T_k of depth d to the quadtree $\oplus_k T_k$ of depth $d+1$ formed by making the T_k the children of a new root node t_0. (Note that $\oplus$ is not commutative since the children are ordered.) Then maximization over the set of all trees of maximum depth D, $\mathcal{T}_D$, of a function $L : \mathcal{T}_D \to \mathbb{R}$ can be written as

$$\max_{T\in\mathcal{T}_D} L^*(T) = \max\left\{ L^*(T_0), \max_{\{T_k\}\in(\mathcal{T}_{D-1})^4} L^*(\oplus_k T_k)\right\} ,$$

where T_0 is the tree with only one node. Using now the fact that $L^*(T) = \sum_{t\in T} E^*(t)$, we have

$$\begin{aligned}\max_{T\in\mathcal{T}_D} L^*(T) &= \max\left\{ L^*(T_0), \max_{\{T_k\}\in(\mathcal{T}_{D-1})^4} \sum_k L^*(T_k)\right\} \\ &= \max\left\{ L^*(T_0), \sum_k \max_{T_k\in\mathcal{T}_{D-1}} L^*(T_k)\right\} .\end{aligned}$$

Coupled with the boundary condition that $\max_{T\in\mathcal{T}_0} L^*(T) = L^*(T_0)$, this equation means that we can use a recursive depth-first search of the tree T_D, the complete quadtree of depth D, to compute the MAP estimate of T, μ, and θ.

4.3 MAP estimates μ^* and θ^*

It remains to specify the functions μ^* and θ^*. These depend only on the first two factors in equation (8), and then only on the subfactors involving node t. From equations (9) and (10), they are thus found by maximizing

$$\Pr(w_t|\theta(t),\mu(t))e^{-\dim(\mu(t))\ln(A)} . \tag{13}$$

The Gaussian model is one-dimensional for all subbands except the scaling coefficients, since it must have zero mean. The scaling coefficient subband is two-dimensional. The generalized Gaussian model is two-dimensional, while the constrained mixture of Gaussians model is four-dimensional.

For each choice of model, $\mu(t) \in M$, the maximizing $\theta(t)$ can be computed from the first factor in the function (13) in ways to be described in sections 4.3.1, 4.3.2, and 4.3.3, and then substituted into the function. The model that maximizes the function, and its corresponding maximizing parameters, are then the values of $\mu^*(t)$ and $\theta^*(t)$.

Note that by following the argument in section 3.1, the distribution $\Pr(w_t|\theta(t),\mu(t))$ gives the histogram of subband t directly; we will therefore use comparisons of the histogram and the distribution to illustrate the fitting of the models.

4.3.1 Gaussian models

For Gaussian models, we have

$$\Pr(w_t|\theta(t),\mu(t)=\mathrm{G}) = (f_t/\pi)^{\frac{N_t}{2}} e^{-f_t \sum_{i\in t}(w_{t,i}-\nu_t)^2} ,$$

where $N_t = |t|$ is the number of wavelet coefficients in the subband t, and $i \in t$ indexes the wavelet packet coefficients within subband t. The parameters of the Gaussian model for each subband are thus f_t, the inverse variance, and ν_t, the mean, which is set equal to zero for all subbands except the scaling coefficients. Then we have that

$$\nu_t^* = \begin{cases} \frac{\sum_{i\in t} w_{t,i}}{N_t} & \text{scaling coefficients ,} \\ 0 & \text{all other subbands ,} \end{cases}$$

$$f_t^* = \frac{N_t}{2\sum_{i\in t}(w_{t,i}-\nu_t^*)^2} .$$

4.3.2 Generalized Gaussian models

For generalized Gaussian models we have that

$$\Pr(w_t|\theta(t),\mu(t)=\mathrm{GG}) = Z^{-N_t}(f_t,s_t)e^{-f_t\sum_{i\in t}|w_{t,i}|^{s_t}} . \tag{14}$$

The number s_t is called the shape factor, and f_t controls the width of the distribution; Z is a normalization constant given by

$$Z = \frac{\Gamma(1+s_t^{-1})}{2f_t^{1/s_t}} .$$

Differentiating the logarithm of equation (14) with respect to the parameters, we find

$$N_t - s_t f_t \sum_{i\in t} |w_{t,i}|^{s_t} = 0$$

$$N_t(\ln(f_t) - \Psi(1+s_t^{-1})) + s_t^2 f_t \sum_{i\in t} |w_{t,i}|^{s_t}\ln(|w_{t,i}|) = 0 \tag{15}$$

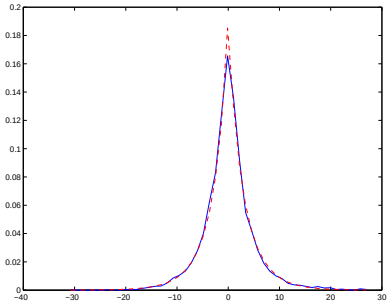


Figure 15: The result of fitting the generalized Gaussian model to the subband t_1 indicated in figure 1. Observed histogram in blue (solid line); fitted model in red (dashed line).

where Ψ is the digamma function. We solve the nonlinear system (15) numerically for f_t and s_t using the algorithm described by Do and Vetterli [11].

Figure 15 shows the fitting of the generalized Gaussian model for the subband t_1 of figure 1.

4.3.3 Constrained mixture of Gaussians model

In order to model the bimodal subbands, we use a constrained mixture of three Gaussians:

$$\begin{aligned}\Pr(w_t|\theta(t),\mu(t)=\text{MoG}) &= \prod_{i\in t}\left[\sum_{a=0}^{2}\frac{P_{t,a}}{(2\pi\sigma_{t,a}^2)^{\frac{1}{2}}}e^{-\frac{(w_{t,i}-\nu_{t,a})^2}{2\sigma_{t,a}^2}}\right]\\ &= \prod_{i\in t}\left[\sum_{a=0}^{2}P_{t,a}G_{t,i}(\nu_{t,a},\sigma_{t,a})\right]\ , \end{aligned} \tag{16}$$

where $a \in \{0,1,2\}$ indexes the mixture components, and the second line defines the functions $G_{t,i}$. The mixture probabilities, $P_{t,a}$, the means $\nu_{t,a}$ and the variances $\sigma_{t,a}^2$ obey the following symmetry constraints:

$$P_{t,1} = P_{t,2} \tag{17}$$

$$\nu_{t,1} = -\nu_{t,2} \tag{18}$$

$$\sigma_{t,1} = \sigma_{t,2} \tag{19}$$

$$\nu_{t,0} = 0\ , \tag{20}$$

in addition to the obvious normalization of the $P_{t,a}$. The mixture of Gaussians model thus has four parameters to be estimated. An example of such a mixture is shown in figure 16.

The parameters can no longer be estimated analytically but, as is well known, the estimation problem can be solved by the expectation-maximization (EM) algorithm [9, 17]. By following the normal derivation of the EM update rules for a mixture of Gaussians, and imposing the constraints in equations (17) to (20) either directly or using Lagrange multipliers, it can be shown that the update equations take the

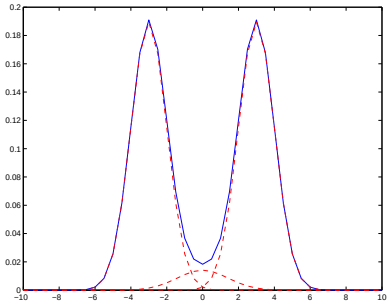


Figure 16: Distributions in bimodal subbands are modelled by a constrained mixture of three Gaussians. An example is shown here. Single components are in red (dashed line), while the mixture, *i.e.* the sum of the three components, is in blue (solid line).

following forms (we have dropped the subscript t indicating the subband for the sake of clarity):

$$P_0' = \frac{1}{N}\sum_i P_0(i) \tag{21}$$

$$\nu_1' = -\nu_2' = \frac{1}{2}\left[\frac{\sum_i w_i P_1(i)}{\sum_i P_1(i)} - \frac{\sum_i w_i P_2(i)}{\sum_i P_2(i)}\right] \tag{22}$$

$$\sigma_1' = \sigma_2' = \left[\frac{\sum_i \sum_{a=1}^{2}(w_i - \nu_a')^2 P_a(i)}{\sum_i \sum_{a=1}^{2} P_a(i)}\right]^{1/2} \tag{23}$$

$$\sigma_0' = \left[\frac{\sum_i w_i^2 P_0(i)}{\sum_i P_0(i)}\right]^{1/2}, \tag{24}$$

where

$$P_a(i) = \frac{G_i(\nu_a, \sigma_a)P_a}{\sum_{a=0}^{2} G_i(\nu_a, \sigma_a)P_a}\,. \tag{25}$$

At each iteration the likelihood of the estimates increases until a local maximum is reached to within a certain tolerance. Although the EM algorithm may converge to a local maximum, its convergence is guaranteed [9, 17].

Figure 17 shows the mixture of Gaussians model trained on the subband t_2 of figure 1.

5 Multimodality: Experiment

In this section, we first discuss the stability of the decompositions with respect to the choice of the parameters A and β. We then present the results of training the model on various textures from the Brodatz album (subsection 5.2) and from remote sensing images (subsection 5.3).

5.1 Stability with respect to A and β

The model described in section 4 requires the selection of two parameters: A, the range of values allowed for the parameters of the various sub-models; and β, the penalty that controls the depth of the partition

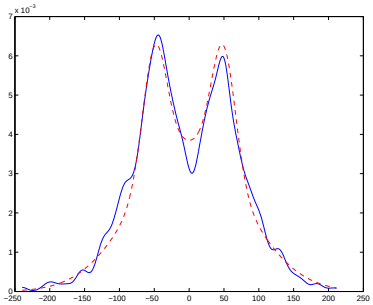


Figure 17: The result of fitting the mixture of Gaussians model to the subband t_2 indicated in figure 1. Observed histogram in blue (solid line); fitted model in red (dashed line).

(see equations (10) and (11) respectively). In the absence of any convincing or tractable method for the elimination or estimation of these parameters, it is necessary to study the stability of the decomposition obtained under changes of these parameters. To this purpose, two different experiments were carried out. In the first experiment, only the β parameter is analysed. In the second experiment, both the A and β parameters are considered.

In order to focus attention on the parameter β, we considered only the Gaussian model. Trials were carried out on different textures for values of β that were multiples of 50 lying in the range $[200, 400]$. In figure 18, the resulting decompositions for the Raffia texture are shown. The decomposition is quite stable and reasonable within this range of values.

In a second experiment, we analysed the stability of the decomposition with respect to both A and β. To this end, values for these parameters were selected at random from a uniform distribution in the ranges $[0, 20]$ and $[50, 500]$. The results obtained for the Raffia texture are shown in figure 19. Despite the different parameter values, the decompositions obtained are very similar. In particular, the multimodal subspace is very nearly the same in all cases.

Although the decompositions are stable within the ranges just described, extreme values do produce unreasonable results. It is thus necessary to tune these parameters to some degree. This can be done during training by running preliminary trials on the texture to be modelled, bearing in mind that β should be proportional to the total amount of training data.

5.2 Modelling Brodatz textures

We trained the new model on a number of different textures taken from the Brodatz album [5]. In all cases, the resulting models match closely the periodicities and histograms of the textures considered, and confirm the link between periodicities and multimodality. Figures 1 and 25 show the results obtained on the Raffia and Herring textures respectively. Similar results where obtained for other textures.

The optimal decomposition of the Raffia texture shown in figure 20 was obtained with $A = 5$ and $\beta = 300$. The decomposition resulted in 57 subbands. Among these, four are modelled by G, 49 by GG, and five by MoG. In figure 21, the observed histograms and the fitted models are shown for each subband. Figure 22 plots the histograms and the fitted models for the bimodal subbands. Some other examples of subband histograms and fitted models are shown in figures 23 and 24.

The optimal decomposition of the Herring texture shown in figure 25 was obtained with the same values of A and β as above. The decomposition resulted in 64 subbands. Among these, nine are modelled by G, 53 by GG, and two by MoG. In figure 26, the observed histograms and the fitted models are shown for each subband. Some other examples of subband histograms and fitted models are shown in figure 27.

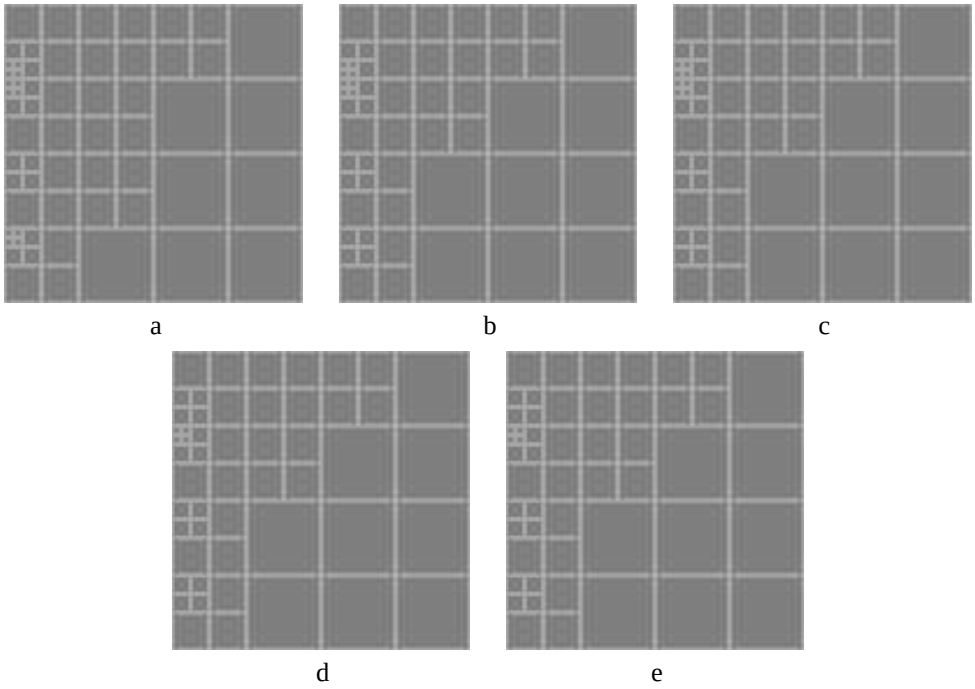


Figure 18: Decompositions obtained for the Raffia texture by varying β: a) $\beta = 200$, b) $\beta = 250$, c) $\beta = 300$, d) $\beta = 350$, e) $\beta = 400$.

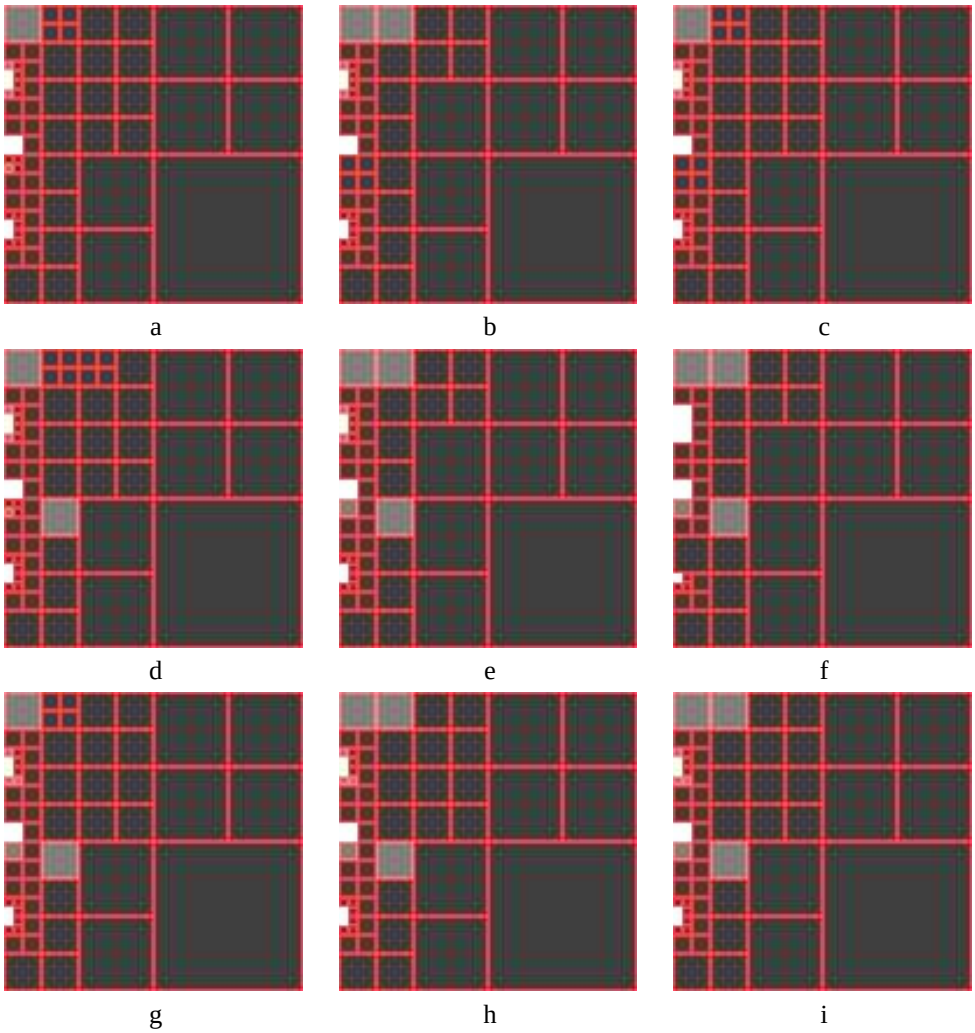
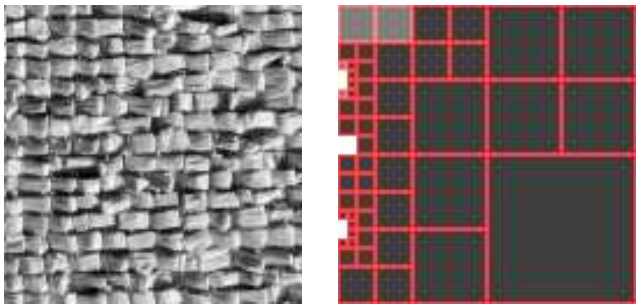


Figure 19: Decompositions obtained for the Raffia texture by varying both A and β. Different colours correspond to the different models automatically selected within each subband: black, grey, and white represent the GG model, the G model, and the MoG model, respectively: a) $A = 4$, $\beta = 97$; b) $A = 5$, $\beta = 200$; c) $A = 9$, $\beta = 153$; d) $A = 11$, $\beta = 76$; e) $A = 12$, $\beta = 249$; f) $A = 13$, $\beta = 409$; g) $A = 14$, $\beta = 152$; h) $A = 14$, $\beta = 171$; i) $A = 14$, $\beta = 190$;



a b

Figure 20: The Raffia texture (a) and its optimal decomposition (b). Different colours correspond to the different models automatically selected within each subband: black, gray and white represent the GG model, the G model, and the MoG model, respectively.

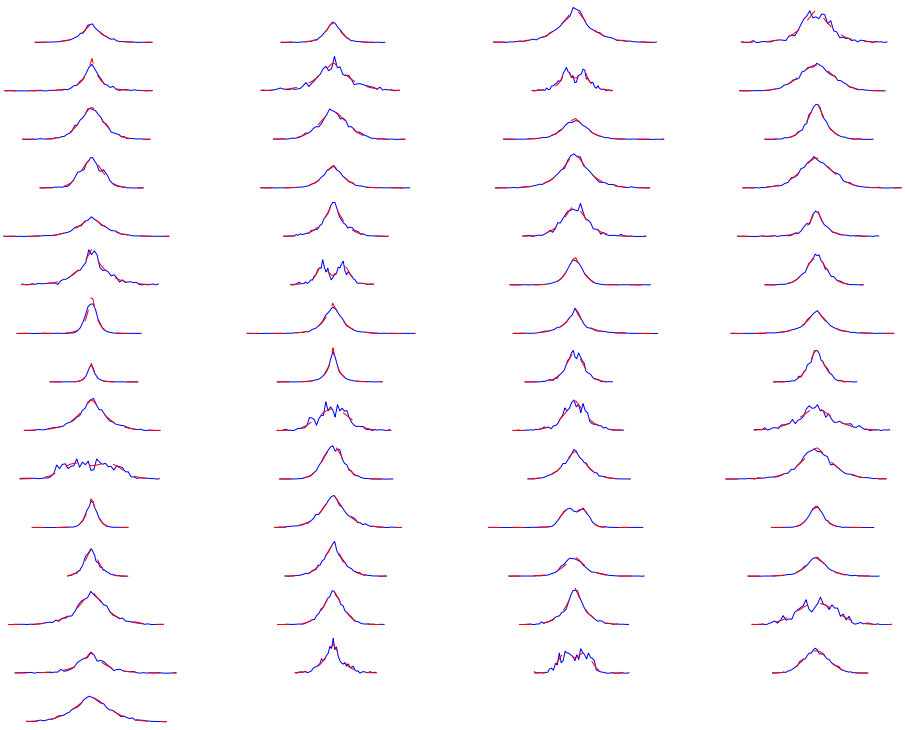


Figure 21: Wavelet coefficient histograms and fitted models for the different subbands into which the Raffia texture is decomposed.

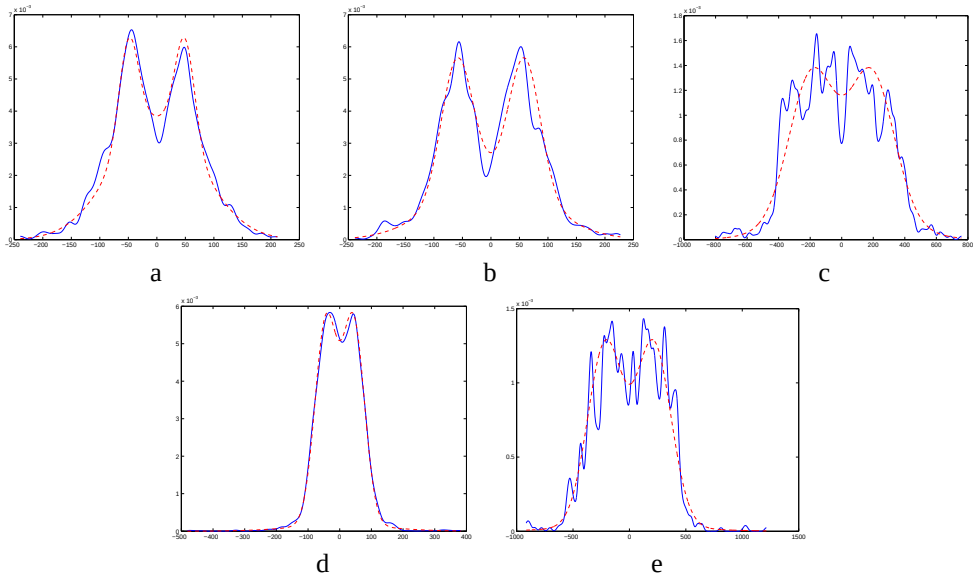


Figure 22: Wavelet coefficient histograms and fitted MoG models for the multimodal subbands in the Raffia texture. Observed histograms in blue (solid line); fitted models in red (dashed line).

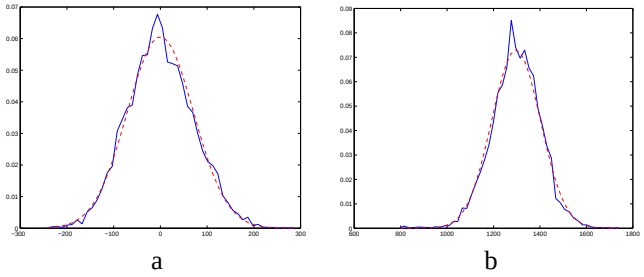


Figure 23: Wavelet coefficient histograms and fitted G models for some subbands in the Raffia texture. The plot on the right corresponds to the scaling coefficients; consequently its mean is different from zero. Observed histograms in blue (solid line); fitted models in red (dashed line).

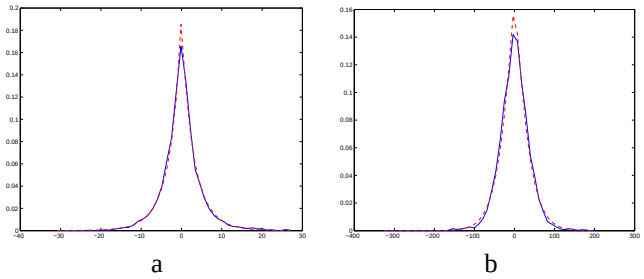
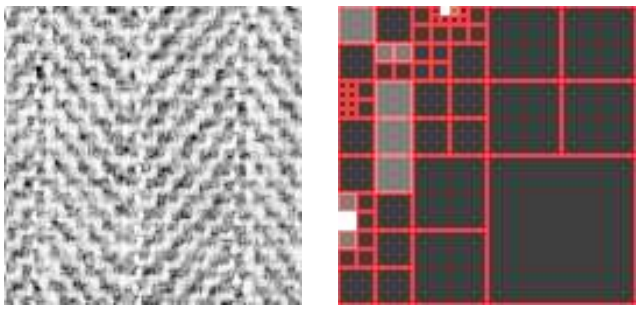


Figure 24: Wavelet coefficient histograms and fitted GG models for some subbands in the Raffia texture. Observed histograms in blue (solid line); fitted models in red (dashed line).



a b

Figure 25: Herring texture (a) and its optimal decomposition (b). Different colours correspond to the different models automatically selected within each subband: black, gray and white represent the GG model, the G model, and the MoG model, respectively.

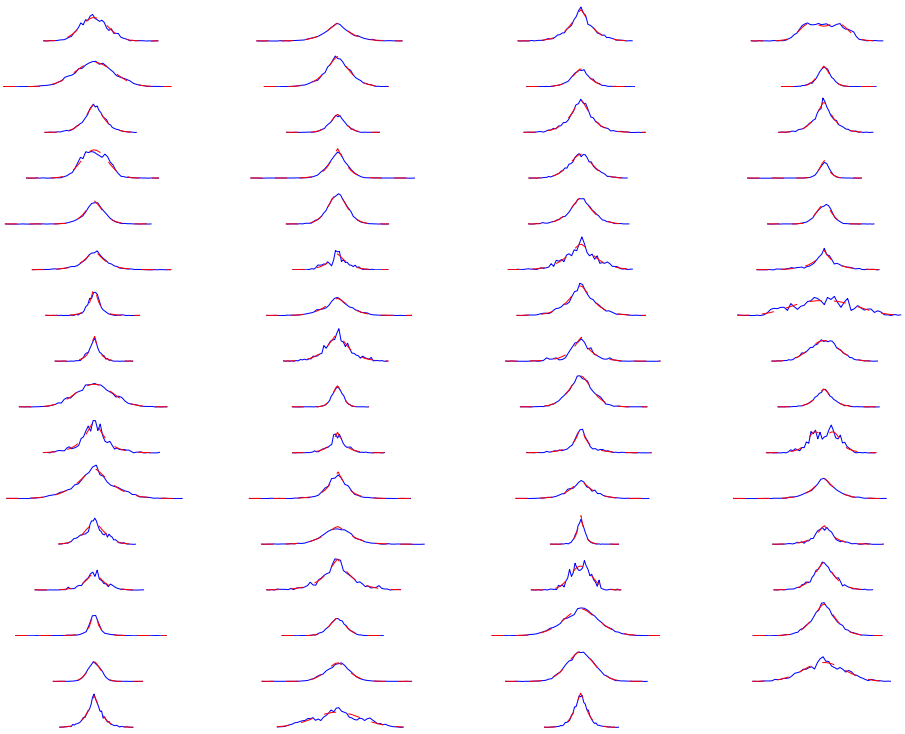


Figure 26: Wavelet coefficient histograms and fitted models for the different subbands in the Herring texture. Observed histograms in blue (solid line); fitted models in red (dashed line).

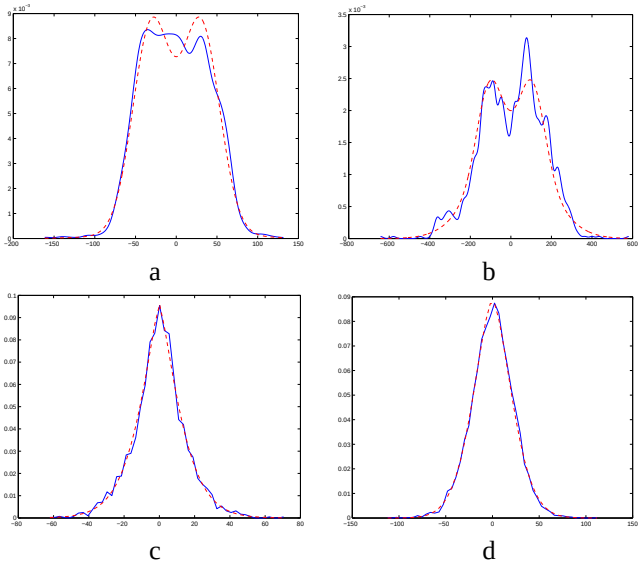


Figure 27: Wavelet coefficient histograms and fitted models for some subbands in the Herring texture. a) to d): MoG, GG, and G model respectively. Observed histograms in blue (solid line); fitted models in red (dashed line).

Figure 28 shows two other examples of Brodatz textures, together with their optimal decompositions, and an example bimodal subband histogram.

5.3 Modelling remote-sensing images

We trained the new model on some textures extracted from remote sensing images. Texture is very important when analysing remote sensing data, especially when high spatial resolution sensors are considered [18].

Figure 29 shows a piece of an aerial image corresponding to a ploughed field. Its optimal decomposition, again obtained with $A = 5$ and $\beta = 300$, is also shown in figure 29. The resulting subband histograms and fitted models are shown in figure 30. There were 54 GG subbands, four G subbands, and nine MoG subbands. The multimodal histograms and fitted models are shown in figure 31.

In figure 32 is shown an image corresponding to an urban area extracted from an aerial image. The texture does not contain any clear periodicities. The decomposition obtained and the subband histograms are plotted in figure 32 and figure 33, respectively. As expected, none of the subbands shows multimodal behaviour.

Figure 34 shows two other examples of remote sensing textures (a forest plantation and a poplar stand), together with their optimal decompositions, and an example bimodal subband histogram.

5.4 The use of bimodal subbands as texture descriptors

For all unimodal subbands, the most probable value of the wavelet packet coefficients is zero, *i.e.* the most probable image composed of these subbands is untextured. Clearly this is unrealistic for a texture model. In contrast, even though the multimodal subbands have zero mean, the most probable value for the coefficients is non-zero, and therefore textured. Since the multimodal subbands typically have narrow

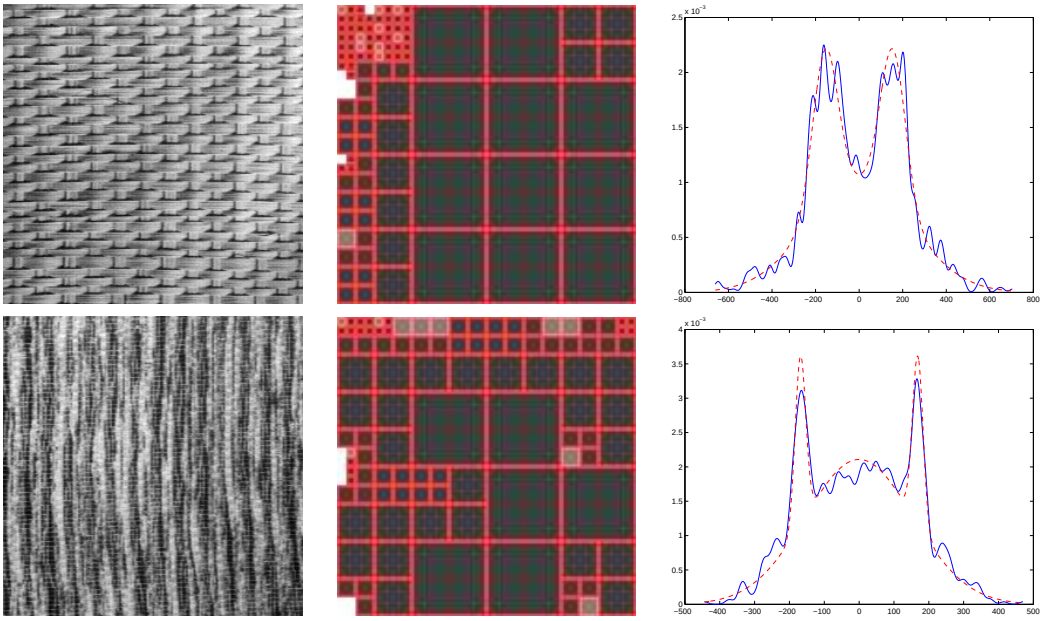
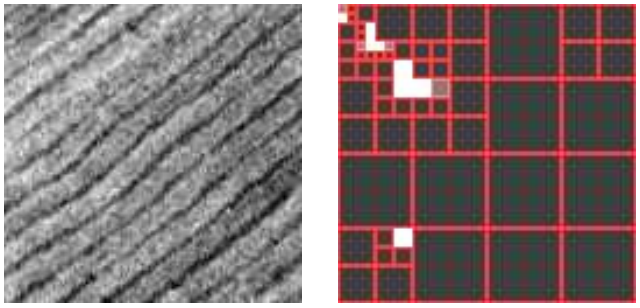


Figure 28: Two Brodatz textures, their optimal decompositions (dark grey = GG, light grey = G, white = MoG), and examples of multimodal subband histograms and the models fitted to them.



a b

Figure 29: Patch extracted from a remote-sensing image and corresponding to a ploughed field (a), and its optimal decomposition (b). Different colours correspond to the different models automatically selected within each subband: black, gray and white represent the GG model, the G model, and the MoG model, respectively.

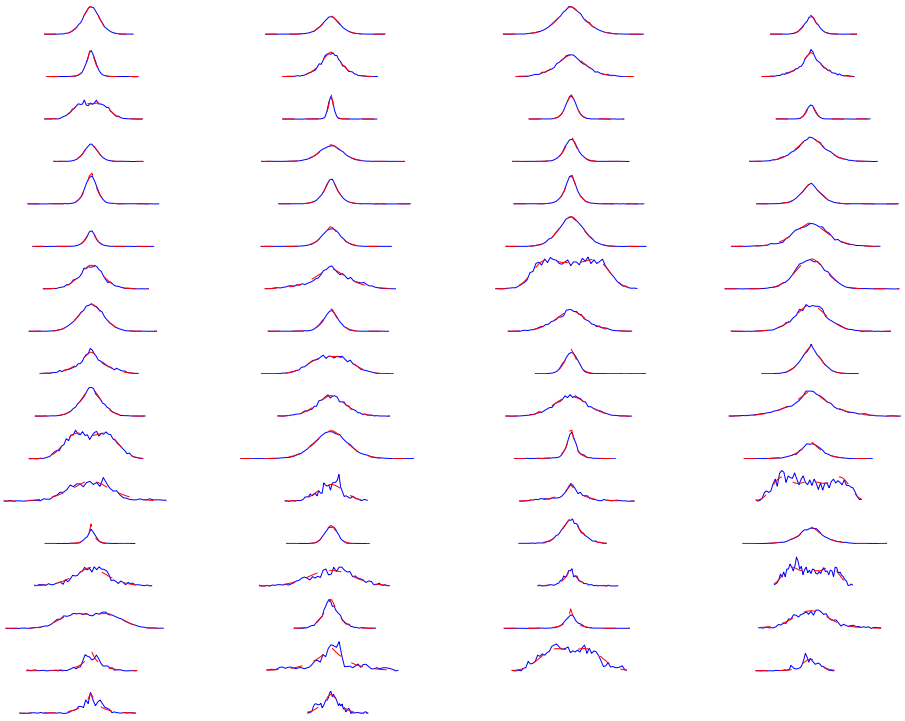


Figure 30: Wavelet coefficient histograms and fitted models for the different subbands in the ploughed field texture. Observed histograms in blue (solid line); fitted models in red (dashed line).

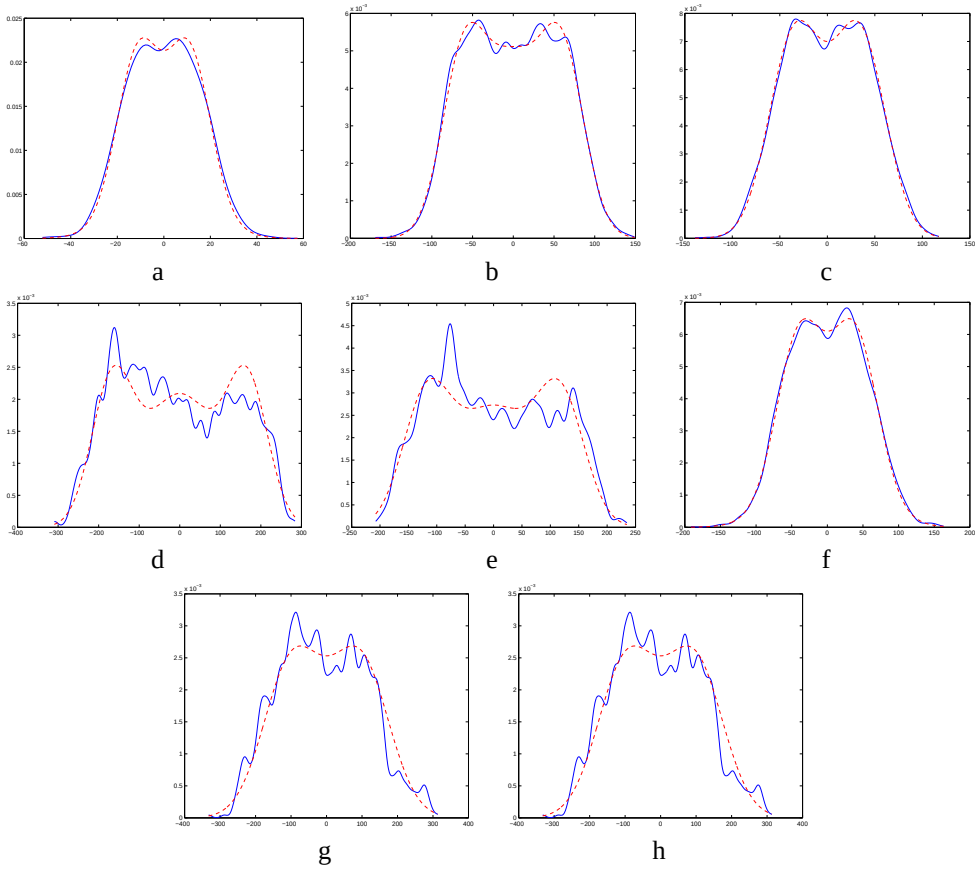


Figure 31: Wavelet coefficient histograms and fitted models for some subbands in the ploughed field texture. Observed histograms in blue (solid line); fitted models in red (dashed line).

Figure 32: Patch extracted from an aerial remote-sensing image and corresponding to an urban area (a), and its optimal decomposition (b). Different colours correspond to the different models automatically selected within each subband: black, gray and white represent the GG model, the G model, and the MoG model, respectively.

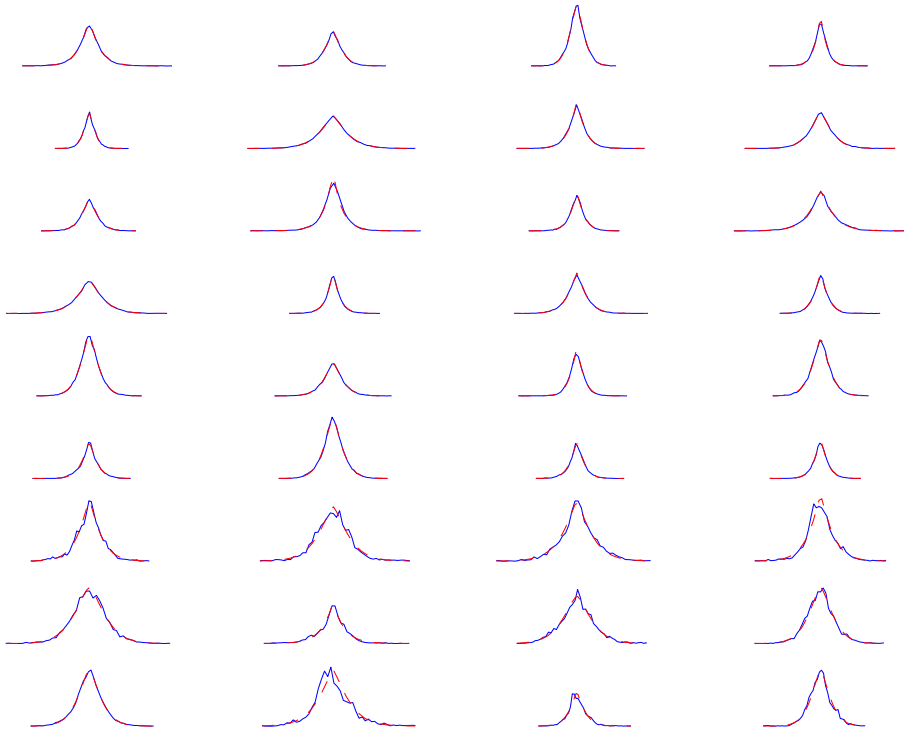


Figure 33: Wavelet coefficient histograms and fitted models for the different subbands in the urban area texture. Observed histograms in blue (solid line); fitted models in red (dashed line).

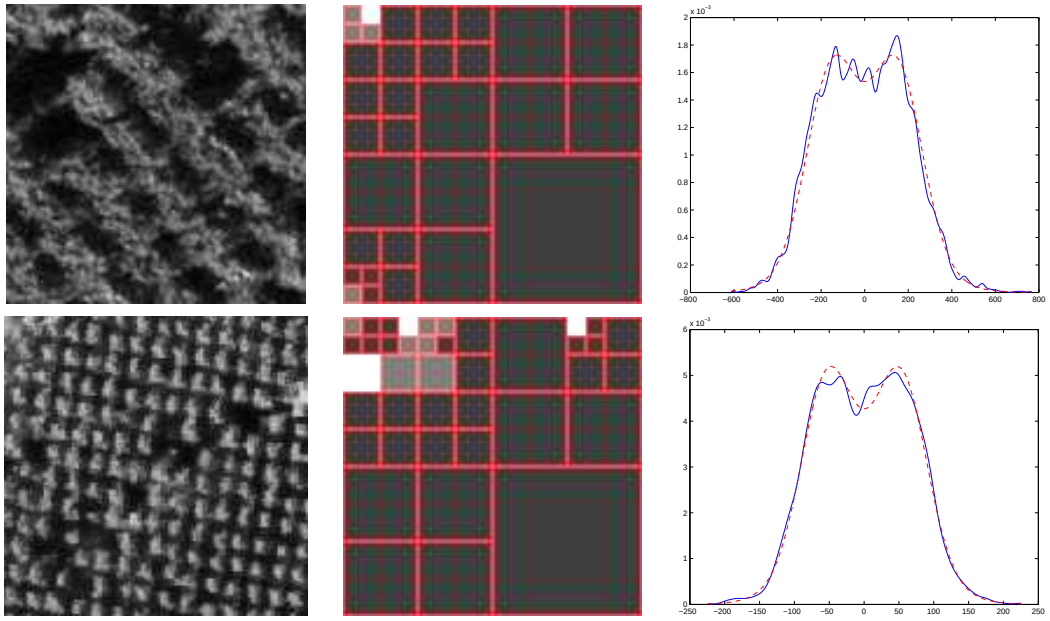


Figure 34: Two remote sensing textures (a forest plantation and a poplar stand), their optimal decompositions (dark grey = GG, light grey = G, white = MoG), and examples of multimodal subband histograms and the models fitted to them.

frequency support, we can think of them as capturing the principal periodicities in the texture, in a manner analogous to the Wold decomposition [12]. Because of their close relation to the structure of the texture, we believe that it is of fundamental importance to identify and model them accurately. For the same reason, multimodal subbands can be used to discriminate between different classes of textures.

Consider again the Raffia (R) and Herring (H) textures. Let $S_i : i \in \{R, H\}$ denote the set of multimodal subbands for each texture. Let $W_{i,j}$ denote the set of $|S_i|$-dimensional vectors of undecimated wavelet packet coefficients from the subbands S_i, computed from texture j. For fixed i, $W_{i,R}$ and $W_{i,H}$ can be plotted as differently coloured points in $|S_i|$ dimensions. The first two rows of figure 35 show projections from such a plot onto two subbands both of which lie in S_R and neither of which lie in S_H. Red/dark patterns correspond to coefficients from Raffia; green/light patterns correspond to coefficients from Herring. As can be seen, the two textures are well separated in this space. The third row shows projections from such a plot onto two subbands one of which is in S_R but not in S_H, the other being in S_H but not in S_R.

For the sake of comparison, figure 36 shows similar plots of two subbands, neither of which lies in either S_R or S_H. The figure shows clearly that for these subbands, the two textures are much less easily separable, the coefficients being mixed together.

Motivated by these plots, in a final series of experiments, we validate the usefulness of bimodal subbands in texture segmentation problems. In all the experiments reported below, the following procedure was used. First, a model was trained for each of the two classes involved. Second, a subband was selected from one of the models, according to criteria that will be detailed in a moment. A new subband model was then trained *for this subband only* for the second class. Third, the undecimated wavelet packet coefficients in this subband were calculated for the original image. Fourth, for each pixel class probabilities were computed by inserting the coefficients belonging to a patch centred on that pixel of size equal to

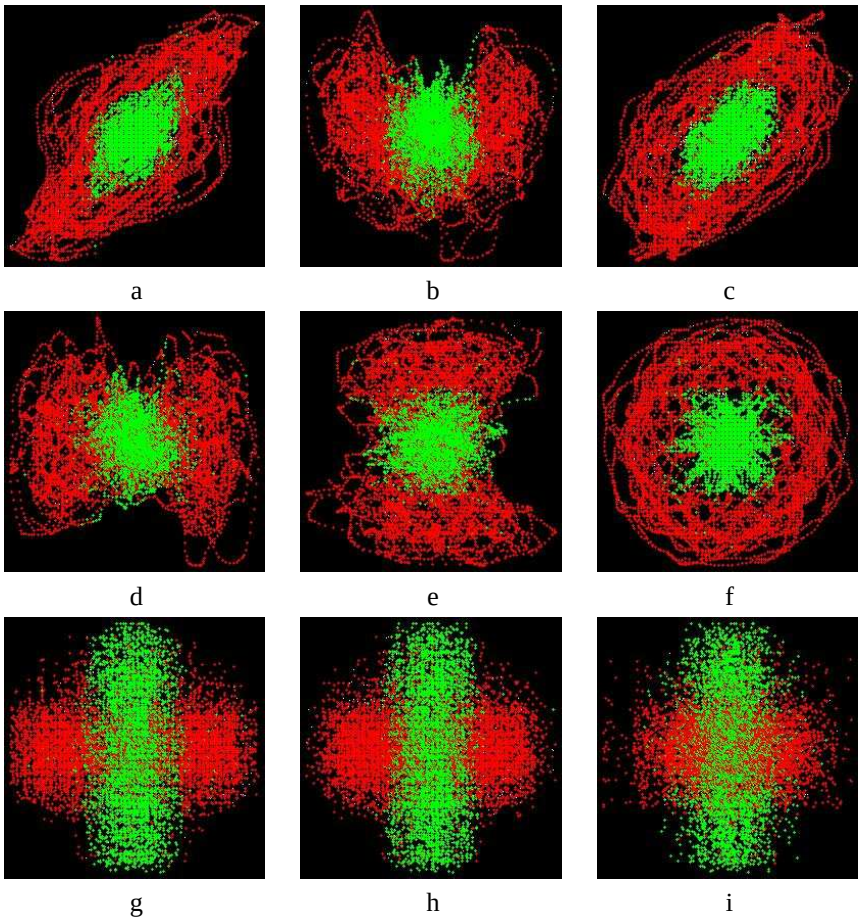
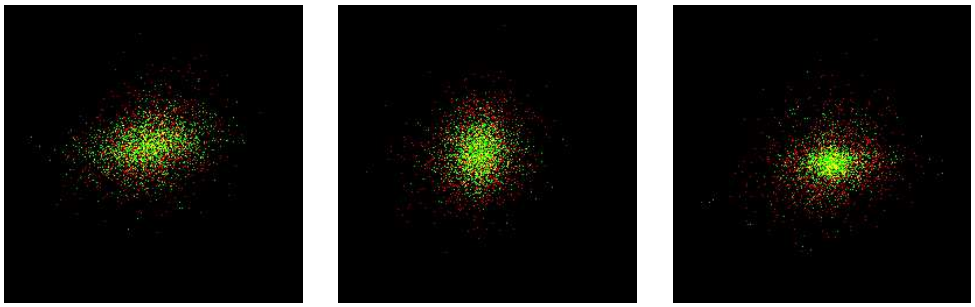


Figure 35: Scatter plots of pairs of undecimated wavelet packet subbands. The first two rows show subbands both of which lie in S_R and neither of which lie in S_H. Red/dark patterns correspond to coefficients from Raffia; green/light patterns correspond to coefficients from Herring. The third row shows subbands one of which is in S_R but not in S_H, the other being in S_H but not in S_R.



a b c

Figure 36: Scatter plots of pairs of undecimated wavelet packet subbands, neither of which lies in either S_R or S_H. Red/dark patterns correspond to coefficients from Raffia; green/light patterns correspond to coefficients from Herring.

the subband filter size into the models for that subband only. Thus only the data lying in this subband were used to compute the probabilities. The differences of the log probabilities of each class form the 'probability map'.

Note that these experiments are not designed to produce the best possible classification results. The full model (*i.e.* using all the subbands) always produces a better result. What is remarkable is that the results of the above procedure using a subband that is multimodal for one texture but not for the other are often very good—indeed for Brodatz textures they approach the performance of the full model—whereas the results using a subband that is unimodal for both textures are very poor.

Figure 37 shows a mosaic made of the Herring and Raffia textures from the Brodatz album. On the right is the result of applying the above procedure using a subband multimodal for Raffia, but unimodal for Herring. Similar results were obtained for other multimodal subbands. The same procedure was then repeated using several subbands of the same size, but unimodal for both textures. The results were all very poor, bearing very little if any connection to the mosaic. Indeed, in many cases the whole image was classified as belonging to one class. The average misclassification rates were as follows. Using Raffia multimodal subbands, the average pixel misclassification rate of Raffia was 14.9%, and of Herring was 0.7%. Using Herring multimodal subbands, Raffia was misclassified 5.1%, and Herring at 17.7%. Using unimodal subbands, the misclassification rate for Herring was 80% and for Raffia 6.5%. The last number is low for the reason that has already been mentioned: most of the unimodal results classified most of the image as Raffia.

Figure 38 shows reconstructions of the mosaic in figure 37 using one of the unimodal subbands (on the left) and one of the multimodal subbands (on the right). The left-hand image shows almost no trace of the mosaic structure, which accounts for the poor classification result. In contrast, the image reconstructed from the multimodal subband clearly shows the mosaic structure, the amplitude of the periodicity being much larger in Raffia than in Herring.

Figure 39 shows a remote sensing image. Models were trained on 'ploughed field' using the upper part of the image, and on 'unploughed field' using the lower part. The above procedure was then followed for a subband multimodal for 'ploughed field' but unimodal for 'unploughed field'. The resulting probability map is shown on the right. The map indicates the ability of the multimodal subband to provide important information for discriminating between the two textures, and in particular to assign a consistently low probability to 'unploughed field'.

Figure 40 shows a second remote sensing image. Models were trained on the 'forest' texture in the lower part of the image, and on a 'background' class consisting of the top-left hand corner. On the right

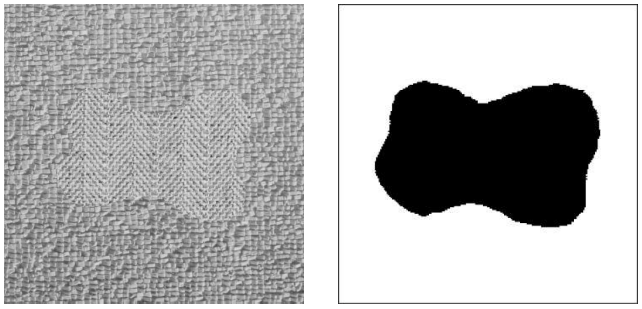


Figure 37: On the left is a mosaic formed of the Herring and Raffia Brodatz textures. On the right, the result of classification using one subband multimodal for Raffia, but unimodal for Herring.

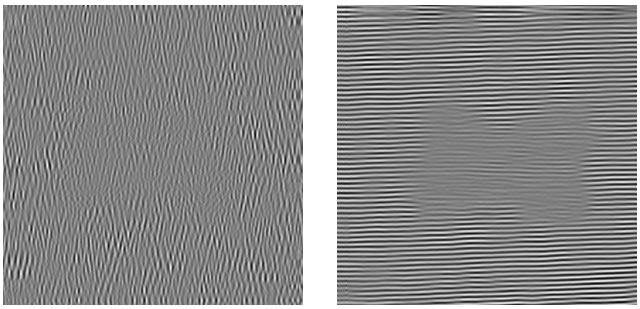


Figure 38: On the left, the reconstruction of the mosaic in figure 37 using a single unimodal subband. On the right, the reconstruction of the mosaic using a single multimodal subband.

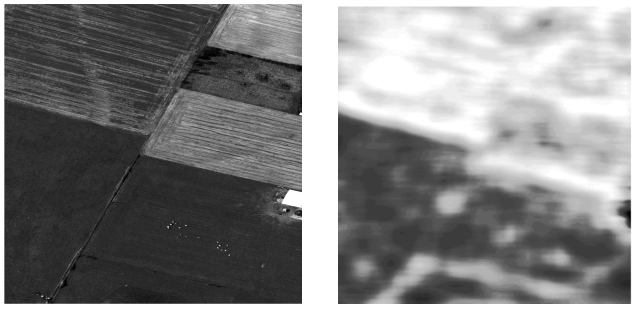


Figure 39: On the left, a remote sensing image. On the right, the probability map resulting from the use of one subband multimodal for the 'ploughed field' texture, but unimodal for the 'unploughed field' texture.

of the figure is shown the result of reconstructing the image using just the multimodal subband in the 'forest' texture. Note how the tree structure of the 'forest' texture has been captured, and how this same structure captures the presence of other trees in the image, while other areas show small response (the image has been linearly stretched for visualization). An exception is the bottom left-hand corner, where the arrangement and size of the buildings represents roughly the same periodicity as the forest. Note

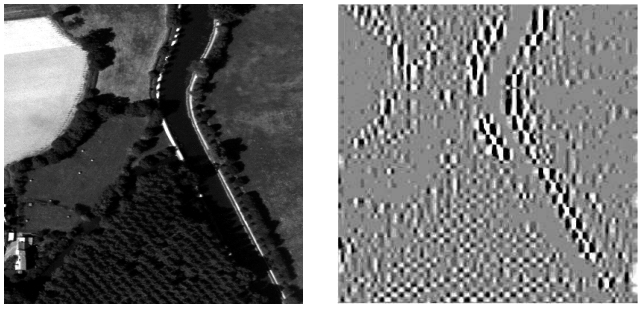


Figure 40: On the left, a remote sensing image. On the right, the reconstruction using a subband multimodal for the ‘forest’ texture in the lower part of the image.

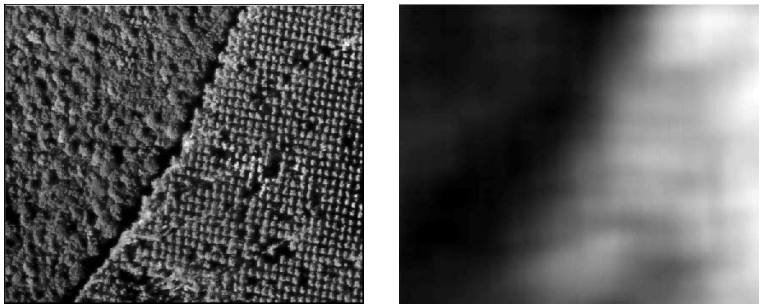


Figure 41: On the left, a remote sensing image. On the right, the probability map resulting from the use of one subband multimodal for the ‘poplar stand’ texture on the right and unimodal for the ‘forest’ texture on the left.

that the subband used here is a standard wavelet subband. Analyzed using a Gaussian or generalized Gaussian model, this subband is not at all remarkable. Nevertheless, the new model was able to detect the multimodality in this subband, and hence capture the structure.

Figure 41 shows another remote sensing image. The classes were ‘poplar stand’, on the right of the image, and ‘forest’ on the left. On the right is shown the probability map resulting from the use of a subband multimodal for ‘poplar stand’, but unimodal for ‘forest’. Note again the consistent assignment of a very low value to ‘forest’, and a much higher value generally to ‘poplar stand’.

Practically speaking, while the classification maps obtained for real remote sensing images are not as accurate as for synthetic images, the probability maps provide useful information. For example, in cases where the classes are spectrally overlapped or otherwise poorly separable, a complex classification system could jointly exploit multispectral information and the information contained in this probability map to obtain a more accurate classification result.

6 Conclusion

Motivated by the observed histograms of adaptive wavelet packet coefficients reported by Brady et al. [2], in this report we have extended the adaptive Bayesian approach of Brady et al. [4] to these statistics by relaxing the Gaussian assumption and allowing three possible models for the wavelet packet coeffi-

cient distributions in each subband of the adaptive basis. We have also described a Bayesian method for estimating the adaptive basis, the subband models, and the subband model parameters. The new model enables the modelling of the newly observed multimodal subbands that capture characteristic structures in the texture. The distributions in these subbands have zero mean, but unlike Gaussian and generalized Gaussian models, the most probable image under these distributions is not textureless. Implicit in these adaptive models are complicated long-range spatial dependencies among the standard wavelet coefficients that it would be hard to capture in any other way.

Experimental results confirm that the models describe the empirical wavelet packet coefficient distributions, and demonstrate the accuracy of the estimation and model selection procedures. They further demonstrate the importance of multimodal subbands for discrimination between and modelling of texture classes. Because of their close relation to structure, it seems probable that the use of multimodal distributions can enable performance improvements in texture classification and denoising. Our current work is studying these problems.

References

[1] J. S. De Bonet and P. A. Viola. A non-parametric multi-scale statistical model for natural images. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10, Denver, USA, 1998. The MIT Press.

[2] K. Brady, I. Jermyn, and J. Zerubia. A probabilistic framework for adaptive texture description. Research Report 4920, INRIA, September 2003. URL `http://www.inria.fr/rrrt/rr-4920.html`.

[3] K. Brady, I. H. Jermyn, and J. Zerubia. Adaptive probabilistic models of wavelet packets for the analysis and segmentation of textured remote sensing images. In *Proc. Brit. Mach. Vis. Conf.*, Norwich, U. K., September 2003.

[4] K. Brady, I. H. Jermyn, and J. Zerubia. Texture analysis: An adaptive probabilistic approach. In *Proc. IEEE Int'l Conf. Im. Proc.*, Barcelona, Spain, September 2003.

[5] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover, New York, USA, 1966.

[6] H. Choi and R. Baraniuk. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Trans. Image Processing*, 10(9):1309–1321, 2001.

[7] R.W. Conners, C.W. McMillin, K. Lin, and R.E. Vasquez Espinosa. Identify and locating surface defects in wood: Part of an automated lumber processing system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:573–583, 1983.

[8] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46(4):886–902, 1998.

[9] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via em algorithm. *Journal of Royal Statistical Society*, 39:1–38, 1977.

[10] P. Dewaele, P. Van Gool, and A. Oosterlinck. Texture inspection with self-adaptive convolution filters. In *Proc. International Conference on Pattern Recognition*, pages 56–60, Rome, Italy, November 14-17 1998.

[11] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Liebler distance. *IEEE Trans. Image Processing*, 11:146–158, 2002. 2.

[12] J. M. Francos, A. Z. Meiri, and B. Porat. A unified texture model based on a 2d Wold-like decomposition. *IEEE Trans. Signal Processing*, 41:2665–2678, 1993.

[13] R. Gupta and P. Undrill. The use of texture analysis to delineate suspicious masses in mammography. *Physics in Medicine and Biology*, 40:835–855, 1995.

[14] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

[15] S. Livens, P. Scheunders, G. Van de Wouwer, and D. Van Dyck. Wavelets for texture analysis, an overview. In *Proc.* 6^{th} *Int'l Conf. Im. Proc. and its Applications*, pages 581–585, Dublin, Ireland, July 1997.

[16] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2^{nd} edition, 1999.

[17] T.K. Moon. The expectation-maximization algorithm. *Signal Processing Magazine*, 13:47–60, 1996.

[18] J. A. Richards. *Remote Sensing Digital Image Analysis*. Springer-Verlag, New York, second edition, 1993.

[19] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice Hall, 1995.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399